

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:27 ; Search time 22.61 Seconds
(without alignments)
40.907 Million cell updates/sec

Title: US-09-439-313-566
Perfect score: 146
Sequence: 1 VGEGLYQGVPRAPETGARRHYDEGVR 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	43.2	742	1 SYV2_RAT	Q02563 rattus norv
2	48	32.9	558	1 DHE3_RAT	P10860 rattus norv
3	47.5	32.5	413	1 INVB_ZYMO	Q60115 zymomonas m
4	47	32.2	457	1 CMGA_HUMAN	P10645 homo sapien
5	47	32.2	558	1 DHE3_HUMAN	P00367 homo sapien
6	46.5	31.8	308	1 ILVE_ECOLI	P00510 escherichia
7	46.5	31.8	917	1 FAN_HUMAN	Q92636 homo sapien
8	46	31.5	249	1 SURE_HAEIN	P45681 haemophilus
9	46	31.5	531	1 LAC4_THACU	Q02081 thanatephor
10	46	31.5	558	1 DHE4_HUMAN	P49448 homo sapien
11	46	31.5	578	1 YBHF_ECOLI	P75776 escherichia
12	46	31.5	615	1 MEN1_HUMAN	O00255 homo sapien
13	46	31.5	637	1 P73_CERAE	Q9xsk8 cercopithec
14	46	31.5	744	1 IE63_HCMVA	P16749 human cytom
15	45.5	31.2	308	1 ILVE_SALTY	P15168 salmonella
16	45	30.8	38	1 EST5_DROMO	P10095 drosophila
17	45	30.8	324	1 FASA_RAT	Q63199 rattus norv
18	45	30.8	557	1 GSHC_TOBAC	P80461 nicotiana t
19	45	30.8	2124	1 PGCA_RAT	P07897 rattus norv
20	44	30.1	217	1 KAD_PARDE	P10772 paracoccus
21	44	30.1	562	1 LCB2_KLULA	P48241 kluyveromyc
22	44	30.1	740	1 PEC1_PIG	Q95242 sus scrofa
23	44	30.1	971	1 VPA1_HALHA	P15350 halobacteri
24	43.5	29.8	153	1 VPG_BYDYN	P27579 barley yell
25	43.5	29.8	305	1 ASGX_PYRAB	Q9x262 pyrococcus
26	43.5	29.8	686	1 RCGG_BORBU	O51528 borrelia bu
27	43.5	29.8	994	1 SYLM_NEUCR	P15181 neurospora
28	43.5	29.8	4544	1 LRP1_HUMAN	Q07954 homo sapien
29	43	29.5	278	1 D7_XENLA	P13007 xenopus lae
30	43	29.5	282	1 CC08_CAEEL	P18833 caenorhabdi
31	43	29.5	381	1 HFLC_ECOLI	P08174 homo sapien
32	42.5	29.1	334	1 MOXR_METEX	P25661 escherichia
33	42.5	29.1	341	1	P30621 methylobact

34	42.5	29.1	357	1 YFOB_SCHPO	Q10170 schizosacch
35	42.5	29.1	389	1 PGK_THETH	P09403 thermus aqu
36	42.5	29.1	445	1 DCDA_CORGL	P09890 corynebacte
37	42	28.8	130	1 ECC1_HALEL	Q9zeu6 halomonas e
38	42	28.8	283	1 YQ33_CAEEL	Q09233 caenorhabdi
39	42	28.8	302	1 CITG_KLEPN	P43414 klebsiella
40	42	28.8	341	1 FLAA_BORBU	P70856 borrelia bu
41	42	28.8	392	1 A2AB_BOVIN	O77700 bos taurus
42	42	28.8	430	1 FTSY_MYCLE	O33010 mycobacteri
43	42	28.8	445	1 CMGA_PIG	P04404 sus scrofa
44	42	28.8	478	1 VPEA_ARATH	P49047 arabidopsis
45	42	28.8	558	1 DHE3_MOUSE	P26443 mus musculu
46	42	28.8	583	1 HAS1_MOUSE	Q61647 mus musculu
47	42	28.8	668	1 BIP2_ARATH	Q39043 arabidopsis
48	42	28.8	694	1 LCF4_YEAST	P47912 saccharomyc
49	42	28.8	699	1 TKT_MYCLE	P46708 mycobacteri
50	42	28.8	700	1 TKT_MYCTU	O08811 mycobacteri
51	42	28.8	844	1 PBPB_ECOLI	P02919 e penicilli
52	42	28.8	953	1 UVRA_RICPR	Q92cc3 rickettsia
53	42	28.8	968	1 ATSL_MOUSE	P97857 mus musculu
54	42	28.8	1224	1 RPOD_PORPU	P51250 porphyra pu
55	42	28.8	1235	1 DNBI_HCMVA	P17147 human cytom
56	42	28.8	1363	1 RPOD_CYAPA	P48120 cyanophora
57	42	28.8	1363	1 VGR3_MOUSE	P35917 mus musculu
58	42	28.8	4969	1 RYNC_RABIT	P30957 oryctolagus
59	41.5	28.4	188	1 GAG_AVEV2	P06937 avian endog
60	41.5	28.4	239	1 GAG_AVEV1	P06936 avian endog
61	41.5	28.4	309	1 GAG_FOJSV	P03326 fujinami sa
62	41.5	28.4	425	1 ERF1_METJA	Q58239 methanococc
63	41.5	28.4	453	1 GAG_AVIMC	P03323 avian myelo
64	41.5	28.4	453	1 GAG_AVIMD	P06444 avian myelo
65	41.5	28.4	490	1 ILVC_ECOLI	P05793 escherichia

ALIGNMENTS

RESULT 1

ID	SVY2_RAT	STANDARD	PRT	742 AA
AC	Q02563;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	SYNAPTIC VESICLE PROTEIN 2 (SV2).			
GN	SV2.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.			
RC	TISSUE=Brain;			
RC	MEDLINE=92390722; PubMed=1519064;			
RX	Bajjalieh S.M., Peterson K., Shingal R., Scheller R.H.;			
RA	"SV2, a brain synaptic vesicle protein homologous to bacterial			
RT	transporters.";			
RL	Science 257:1271-1273(1992).			
CC	-1- FUNCTION: MAY ACT AS A VESICLE-LOCALIZED NEUROTRANSMITTER			
CC	TRANSPORTER OR MAY FUNCTION AS AN ION TRANSPORTER OR CHANNEL.			
CC	-1- SUBCELLULAR LOCATION: SYNAPTIC VESICLE.			
CC	-1- TISSUE SPECIFICITY: NEURAL AND ENDOCRINE CELLS OF BRAIN AND			
CC	SPINAL CORD.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; L05435; AAA42188.1; -.			

DR Pfam: PF00083; sugar_tr; 1;
KW Synapse; Nerve; Glycoprotein; Neurotransmitter transport;
FT DOMAIN 1 163
FT TRANSMEM 164 182
FT TRANSMEM 183 204
FT TRANSMEM 205 225
FT TRANSMEM 226 242
FT TRANSMEM 243 262
FT TRANSMEM 263 281
FT TRANSMEM 282 294
FT TRANSMEM 295 322
FT TRANSMEM 323 334
FT TRANSMEM 335 352
FT TRANSMEM 353 370
FT TRANSMEM 371 388
FT TRANSMEM 389 406
FT TRANSMEM 407 424
FT TRANSMEM 425 442
FT TRANSMEM 443 460
FT TRANSMEM 461 478
FT TRANSMEM 479 496
FT TRANSMEM 497 514
FT TRANSMEM 515 532
FT TRANSMEM 533 550
FT TRANSMEM 551 568
FT TRANSMEM 569 586
FT TRANSMEM 587 604
FT TRANSMEM 605 622
FT TRANSMEM 623 640
FT TRANSMEM 641 658
FT TRANSMEM 659 676
FT TRANSMEM 677 694
FT TRANSMEM 695 712
FT TRANSMEM 713 730
FT TRANSMEM 731 748
FT TRANSMEM 749 766
FT TRANSMEM 767 784
FT TRANSMEM 785 802
FT TRANSMEM 803 820
FT TRANSMEM 821 838
FT TRANSMEM 839 856
FT TRANSMEM 857 874
FT TRANSMEM 875 892
FT TRANSMEM 893 910
FT TRANSMEM 911 928
FT TRANSMEM 929 946
FT TRANSMEM 947 964
FT TRANSMEM 965 982
FT TRANSMEM 983 1000
FT TRANSMEM 1001 1018
FT TRANSMEM 1019 1036
FT TRANSMEM 1037 1054
FT TRANSMEM 1055 1072
FT TRANSMEM 1073 1090
FT TRANSMEM 1091 1108
FT TRANSMEM 1109 1126
FT TRANSMEM 1127 1144
FT TRANSMEM 1145 1162
FT TRANSMEM 1163 1180
FT TRANSMEM 1181 1198
FT TRANSMEM 1199 1216
FT TRANSMEM 1217 1234
FT TRANSMEM 1235 1252
FT TRANSMEM 1253 1270
FT TRANSMEM 1271 1288
FT TRANSMEM 1289 1306
FT TRANSMEM 1307 1324
FT TRANSMEM 1325 1342
FT TRANSMEM 1343 1360
FT TRANSMEM 1361 1378
FT TRANSMEM 1379 1396
FT TRANSMEM 1397 1414
FT TRANSMEM 1415 1432
FT TRANSMEM 1433 1450
FT TRANSMEM 1451 1468
FT TRANSMEM 1469 1486
FT TRANSMEM 1487 1504
FT TRANSMEM 1505 1522
FT TRANSMEM 1523 1540
FT TRANSMEM 1541 1558
FT TRANSMEM 1559 1576
FT TRANSMEM 1577 1594
FT TRANSMEM 1595 1612
FT TRANSMEM 1613 1630
FT TRANSMEM 1631 1648
FT TRANSMEM 1649 1666
FT TRANSMEM 1667 1684
FT TRANSMEM 1685 1702
FT TRANSMEM 1703 1720
FT TRANSMEM 1721 1738
FT TRANSMEM 1739 1756
FT TRANSMEM 1757 1774
FT TRANSMEM 1775 1792
FT TRANSMEM 1793 1810
FT TRANSMEM 1811 1828
FT TRANSMEM 1829 1846
FT TRANSMEM 1847 1864
FT TRANSMEM 1865 1882
FT TRANSMEM 1883 1900
FT TRANSMEM 1901 1918
FT TRANSMEM 1919 1936
FT TRANSMEM 1937 1954
FT TRANSMEM 1955 1972
FT TRANSMEM 1973 1990
FT TRANSMEM 1991 2008
FT TRANSMEM 2009 2026
FT TRANSMEM 2027 2044
FT TRANSMEM 2045 2062
FT TRANSMEM 2063 2080
FT TRANSMEM 2081 2098
FT TRANSMEM 2099 2116
FT TRANSMEM 2117 2134
FT TRANSMEM 2135 2152
FT TRANSMEM 2153 2170
FT TRANSMEM 2171 2188
FT TRANSMEM 2189 2206
FT TRANSMEM 2207 2224
FT TRANSMEM 2225 2242
FT TRANSMEM 2243 2260
FT TRANSMEM 2261 2278
FT TRANSMEM 2279 2296
FT TRANSMEM 2297 2314
FT TRANSMEM 2315 2332
FT TRANSMEM 2333 2350
FT TRANSMEM 2351 2368
FT TRANSMEM 2369 2386
FT TRANSMEM 2387 2404
FT TRANSMEM 2405 2422
FT TRANSMEM 2423 2440
FT TRANSMEM 2441 2458
FT TRANSMEM 2459 2476
FT TRANSMEM 2477 2494
FT TRANSMEM 2495 2512
FT TRANSMEM 2513 2530
FT TRANSMEM 2531 2548
FT TRANSMEM 2549 2566
FT TRANSMEM 2567 2584
FT TRANSMEM 2585 2602
FT TRANSMEM 2603 2620
FT TRANSMEM 2621 2638
FT TRANSMEM 2639 2656
FT TRANSMEM 2657 2674
FT TRANSMEM 2675 2692
FT TRANSMEM 2693 2710
FT TRANSMEM 2711 2728
FT TRANSMEM 2729 2746
FT TRANSMEM 2747 2764
FT TRANSMEM 2765 2782
FT TRANSMEM 2783 2800
FT TRANSMEM 2801 2818
FT TRANSMEM 2819 2836
FT TRANSMEM 2837 2854
FT TRANSMEM 2855 2872
FT TRANSMEM 2873 2890
FT TRANSMEM 2891 2908
FT TRANSMEM 2909 2926
FT TRANSMEM 2927 2944
FT TRANSMEM 2945 2962
FT TRANSMEM 2963 2980
FT TRANSMEM 2981 2998
FT TRANSMEM 2999 3016
FT TRANSMEM 3017 3034
FT TRANSMEM 3035 3052
FT TRANSMEM 3053 3070
FT TRANSMEM 3071 3088
FT TRANSMEM 3089 3106
FT TRANSMEM 3107 3124
FT TRANSMEM 3125 3142
FT TRANSMEM 3143 3160
FT TRANSMEM 3161 3178
FT TRANSMEM 3179 3196
FT TRANSMEM 3197 3214
FT TRANSMEM 3215 3232
FT TRANSMEM 3233 3250
FT TRANSMEM 3251 3268
FT TRANSMEM 3269 3286
FT TRANSMEM 3287 3304
FT TRANSMEM 3305 3322
FT TRANSMEM 3323 3340
FT TRANSMEM 3341 3358
FT TRANSMEM 3359 3376
FT TRANSMEM 3377 3394
FT TRANSMEM 3395 3412
FT TRANSMEM 3413 3430
FT TRANSMEM 3431 3448
FT TRANSMEM 3449 3466
FT TRANSMEM 3467 3484
FT TRANSMEM 3485 3502
FT TRANSMEM 3503 3520
FT TRANSMEM 3521 3538
FT TRANSMEM 3539 3556
FT TRANSMEM 3557 3574
FT TRANSMEM 3575 3592
FT TRANSMEM 3593 3610
FT TRANSMEM 3611 3628
FT TRANSMEM 3629 3646
FT TRANSMEM 3647 3664
FT TRANSMEM 3665 3682
FT TRANSMEM 3683 3700
FT TRANSMEM 3701 3718
FT TRANSMEM 3719 3736
FT TRANSMEM 3737 3754
FT TRANSMEM 3755 3772
FT TRANSMEM 3773 3790
FT TRANSMEM 3791 3808
FT TRANSMEM 3809 3826
FT TRANSMEM 3827 3844
FT TRANSMEM 3845 3862
FT TRANSMEM 3863 3880
FT TRANSMEM 3881 3898
FT TRANSMEM 3899 3916
FT TRANSMEM 3917 3934
FT TRANSMEM 3935 3952
FT TRANSMEM 3953 3970
FT TRANSMEM 3971 3988
FT TRANSMEM 3989 4006
FT TRANSMEM 4007 4024
FT TRANSMEM 4025 4042
FT TRANSMEM 4043 4060
FT TRANSMEM 4061 4078
FT TRANSMEM 4079 4096
FT TRANSMEM 4097 4114
FT TRANSMEM 4115 4132
FT TRANSMEM 4133 4150
FT TRANSMEM 4151 4168
FT TRANSMEM 4169 4186
FT TRANSMEM 4187 4204
FT TRANSMEM 4205 4222
FT TRANSMEM 4223 4240
FT TRANSMEM 4241 4258
FT TRANSMEM 4259 4276
FT TRANSMEM 4277 4294
FT TRANSMEM 4295 4312
FT TRANSMEM 4313 4330
FT TRANSMEM 4331 4348
FT TRANSMEM 4349 4366
FT TRANSMEM 4367 4384
FT TRANSMEM 4385 4402
FT TRANSMEM 4403 4420
FT TRANSMEM 4421 4438
FT TRANSMEM 4439 4456
FT TRANSMEM 4457 4474
FT TRANSMEM 4475 4492
FT TRANSMEM 4493 4510
FT TRANSMEM 4511 4528
FT TRANSMEM 4529 4546
FT TRANSMEM 4547 4564
FT TRANSMEM 4565 4582
FT TRANSMEM 4583 4600
FT TRANSMEM 4601 4618
FT TRANSMEM 4619 4636
FT TRANSMEM 4637 4654
FT TRANSMEM 4655 4672
FT TRANSMEM 4673 4690
FT TRANSMEM 4691 4708
FT TRANSMEM 4709 4726
FT TRANSMEM 4727 4744
FT TRANSMEM 4745 4762
FT TRANSMEM 4763 4780
FT TRANSMEM 4781 4798
FT TRANSMEM 4799 4816
FT TRANSMEM 4817 4834
FT TRANSMEM 4835 4852
FT TRANSMEM 4853 4870
FT TRANSMEM 4871 4888
FT TRANSMEM 4889 4906
FT TRANSMEM 4907 4924
FT TRANSMEM 4925 4942
FT TRANSMEM 4943 4960
FT TRANSMEM 4961 4978
FT TRANSMEM 4979 4996
FT TRANSMEM 4997 5014
FT TRANSMEM 5015 5032
FT TRANSMEM 5033 5050
FT TRANSMEM 5051 5068
FT TRANSMEM 5069 5086
FT TRANSMEM 5087 5104
FT TRANSMEM 5105 5122
FT TRANSMEM 5123 5140
FT TRANSMEM 5141 5158
FT TRANSMEM 5159 5176
FT TRANSMEM 5177 5194
FT TRANSMEM 5195 5212
FT TRANSMEM 5213 5230
FT TRANSMEM 5231 5248
FT TRANSMEM 5249 5266
FT TRANSMEM 5267 5284
FT TRANSMEM 5285 5302
FT TRANSMEM 5303 5320
FT TRANSMEM 5321 5338
FT TRANSMEM 5339 5356
FT TRANSMEM 5357 5374
FT TRANSMEM 5375 5392
FT TRANSMEM 5393 5410
FT TRANSMEM 5411 5428
FT TRANSMEM 5429 5446
FT TRANSMEM 5447 5464
FT TRANSMEM 5465 5482
FT TRANSMEM 5483 5500
FT TRANSMEM 5501 5518
FT TRANSMEM 5519 5536
FT TRANSMEM 5537 5554
FT TRANSMEM 5555 5572
FT TRANSMEM 5573 5590
FT TRANSMEM 5591 5608
FT TRANSMEM 5609 5626
FT TRANSMEM 5627 5644
FT TRANSMEM 5645 5662
FT TRANSMEM 5663 5680
FT TRANSMEM 5681 5698
FT TRANSMEM 5699 5716
FT TRANSMEM 5717 5734
FT TRANSMEM 5735 5752
FT TRANSMEM 5753 5770
FT TRANSMEM 5771 5788
FT TRANSMEM 5789 5806
FT TRANSMEM 5807 5824
FT TRANSMEM 5825 5842
FT TRANSMEM 5843 5860
FT TRANSMEM 5861 5878
FT TRANSMEM 5879 5896
FT TRANSMEM 5897 5914
FT TRANSMEM 5915 5932
FT TRANSMEM 5933 5950
FT TRANSMEM 5951 5968
FT TRANSMEM 5969 5986
FT TRANSMEM 5987 6004
FT TRANSMEM 6005 6022
FT TRANSMEM 6023 6040
FT TRANSMEM 6041 6058
FT TRANSMEM 6059 6076
FT TRANSMEM 6077 6094
FT TRANSMEM 6095 6112
FT TRANSMEM 6113 6130
FT TRANSMEM 6131 6148
FT TRANSMEM 6149 6166
FT TRANSMEM 6167 6184
FT TRANSMEM 6185 6202
FT TRANSMEM 6203 6220
FT TRANSMEM 6221 6238
FT TRANSMEM 6239 6256
FT TRANSMEM 6257 6274
FT TRANSMEM 6275 6292
FT TRANSMEM 6293 6310
FT TRANSMEM 6311 6328
FT TRANSMEM 6329 6346
FT TRANSMEM 6347 6364
FT TRANSMEM 6365 6382
FT TRANSMEM 6383 6400
FT TRANSMEM 6401 6418
FT TRANSMEM 6419 6436
FT TRANSMEM 6437 6454
FT TRANSMEM 6455 6472
FT TRANSMEM 6473 6490
FT TRANSMEM 6491 6508
FT TRANSMEM 6509 6526
FT TRANSMEM 6527 6544
FT TRANSMEM 6545 6562
FT TRANSMEM 6563 6580
FT TRANSMEM 6581 6598
FT TRANSMEM 6599 6616
FT TRANSMEM 6617 6634
FT TRANSMEM 6635 6652
FT TRANSMEM 6653 6670
FT TRANSMEM 6671 6688
FT TRANSMEM 6689 6706
FT TRANSMEM 6707 6724
FT TRANSMEM 6725 6742
FT TRANSMEM 6743 6760
FT TRANSMEM 6761 6778
FT TRANSMEM 6779 6796
FT TRANSMEM 6797 6814
FT TRANSMEM 6815 6832
FT TRANSMEM 6833 6850
FT TRANSMEM 6851 6868
FT TRANSMEM 6869 6886
FT TRANSMEM 6887 6904
FT TRANSMEM 6905 6922
FT TRANSMEM 6923 6940
FT TRANSMEM 6941 6958
FT TRANSMEM 6959 6976
FT TRANSMEM 6977 6994
FT TRANSMEM 6995 7012
FT TRANSMEM 7013 7030
FT TRANSMEM 7031 7048
FT TRANSMEM 7049 7066
FT TRANSMEM 7067 7084
FT TRANSMEM 7085 7102
FT TRANSMEM 7103 7120
FT TRANSMEM 7121 7138
FT TRANSMEM 7139 7156
FT TRANSMEM 7157 7174
FT TRANSMEM 7175 7192
FT TRANSMEM 7193 7210
FT TRANSMEM 7211 7228
FT TRANSMEM 7229 7246
FT TRANSMEM 7247 7264
FT TRANSMEM 7265 7282
FT TRANSMEM 7283 7300
FT TRANSMEM 7301 7318
FT TRANSMEM 7319 7336
FT TRANSMEM 7337 7354
FT TRANSMEM 7355 7372
FT TRANSMEM 7373 7390
FT TRANSMEM 7391 7408
FT TRANSMEM 7409 7426
FT TRANSMEM 7427 7444
FT TRANSMEM 7445 7462
FT TRANSMEM 7463 7480
FT TRANSMEM 7481 7498
FT TRANSMEM 7499 7516
FT TRANSMEM 7517 7534
FT TRANSMEM 7535 7552
FT TRANSMEM 7553 7570
FT TRANSMEM 7571 7588
FT TRANSMEM 7589 7606
FT TRANSMEM 7607 7624
FT TRANSMEM 7625 7642
FT TRANSMEM 7643 7660
FT TRANSMEM 7661 7678
FT TRANSMEM 7679 7696
FT TRANSMEM 7697 7714
FT TRANSMEM 7715 7732
FT TRANSMEM 7733 7750
FT TRANSMEM 7751 7768
FT TRANSMEM 7769 7786
FT TRANSMEM 7787 7804
FT TRANSMEM 7805 7822
FT TRANSMEM 7823 7840
FT TRANSMEM 7841 7858
FT TRANSMEM 7859 7876
FT TRANSMEM 7877 7894
FT TRANSMEM 7895 7912
FT TRANSMEM 7913 7930
FT TRANSMEM 7931 7948
FT TRANSMEM 7949 7966
FT TRANSMEM 7967 7984
FT TRANSMEM 7985 8002
FT TRANSMEM 8003 8020
FT TRANSMEM 8021 8038
FT TRANSMEM 8039 8056
FT TRANSMEM 8057 8074
FT TRANSMEM 8075 8092
FT TRANSMEM 8093 8110
FT TRANSMEM 8111 8128
FT TRANSMEM 8129 8146
FT TRANSMEM 8147 8164
FT TRANSMEM 8165 8182
FT TRANSMEM 8183 8200
FT TRANSMEM 8201 8218
FT TRANSMEM 8219 8236
FT TRANSMEM 8237 8254
FT TRANSMEM 8255 8272
FT TRANSMEM 8273 8290
FT TRANSMEM 8291 8308
FT TRANSMEM 8309 8326
FT TRANSMEM 8327 8344
FT TRANSMEM 8345 8362
FT TRANSMEM 8363 8380
FT TRANSMEM 8381 8398
FT TRANSMEM 8399 8416
FT TRANSMEM 8417 8434
FT TRANSMEM 8435 8452
FT TRANSMEM 8453 8470
FT TRANSMEM 8471 8488
FT TRANSMEM 8489 8506
FT TRANSMEM 8507 8524
FT TRANSMEM 8525 8542
FT TRANSMEM 8543 8560
FT TRANSMEM 8561 8578
FT TRANSMEM 8579 8596
FT TRANSMEM 8597 8614
FT TRANSMEM 8615 8632
FT TRANSMEM 8633 8650
FT TRANSMEM 8651 8668
FT TRANSMEM 8669 8686
FT TRANSMEM 8687 8704
FT TRANSMEM 8705 8722
FT TRANSMEM 8723 8740
FT TRANSMEM 8741 8758
FT TRANSMEM 8759 8776
FT TRANSMEM 8777 8794
FT TRANSMEM 8795 8812
FT TRANSMEM 8813 8830
FT TRANSMEM 8831 8848
FT TRANSMEM 8849 8866
FT TRANSMEM 8867 8884
FT TRANSMEM 8885 8902
FT TRANSMEM 8903 8920
FT TRANSMEM 8921 8938
FT TRANSMEM 8939 8956
FT TRANSMEM 8957 8974
FT TRANSMEM 8975 8992
FT TRANSMEM 8993 9010
FT TRANSMEM 9011 9028
FT TRANSMEM 9029 9046
FT TRANSMEM 9047 9064
FT TRANSMEM 9065 9082
FT TRANSMEM 9083 9100
FT TRANSMEM 9101 9118
FT TRANSMEM 9119 9136
FT TRANSMEM 9137 9154
FT TRANSMEM 9155 9172
FT TRANSMEM 9173 9190
FT TRANSMEM 9191 9208
FT TRANSMEM 9209 9226
FT TRANSMEM 9227 9244
FT TRANSMEM 9245 9262
FT TRANSMEM 9263 9280
FT TRANSMEM 9281 9298
FT TRANSMEM 9299 9316
FT TRANSMEM 9317 9334
FT TRANSMEM 9335 9352
FT TRANSMEM 9353 9370
FT TRANSMEM 9371 9388
FT TRANSMEM 9389 9406
FT TRANSMEM 9407 9424
FT TRANSMEM 9425 9442
FT TRANSMEM 9443 9460
FT TRANSMEM 9461 9478
FT TRANSMEM 9479 9496
FT TRANSMEM 9497 9514
FT TRANSMEM 9515 9532
FT TRANSMEM 9533 9550
FT TRANSMEM 9551 9568
FT TRANSMEM 9569 9586
FT TRANSMEM 9587 9604
FT TRANSMEM 9605 9622
FT TRANSMEM 9623 9640
FT TRANSMEM 9641 9658
FT TRANSMEM 9659 9676
FT TRANSMEM 9677 9694
FT TRANSMEM 9695 9712
FT TRANSMEM 9713 9730
FT TRANSMEM 9731 9748
FT TRANSMEM 9749 9766
FT TRANSMEM 9767 9784
FT TRANSMEM 9785 9802
FT TRANSMEM 9803 9820
FT TRANSMEM 9821 9838
FT TRANSMEM 9839 9856
FT TRANSMEM 9857 9874
FT TRANSMEM 9875 9892
FT TRANSMEM 9893 9910
FT TRANSMEM 9911 9928
FT TRANSMEM 9929 9946
FT TRANSMEM 9947 9964
FT TRANSMEM 9965 9982
FT TRANSMEM 9983 10000

Query Match 43.2%; Score 63; DB 1; Length 742;
Best Local Similarity 51.7%; Pred. No. 22; Mismatches 8; Indels 4; Gaps 1;
Matches 15; Conservative 2; Mismatches 8; Indels 4; Gaps 1;
OY 3 EGYGVPRAPETGTEARRHYD---EGVR 27
DB 95 EGEYQIPRAESGGKGERMADGAPLAGVR 123
RESULT 2
ID INVB_ZYMO STANDARD; PRT; 558 AA.
AC P10860;
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE 01-JUL-1989 (Rel. 11, Last sequence update)
DE GLUTAMATE DEHYDROGENASE PRECURSOR (EC 1.4.1.3) (GDH).
GN Glutinus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=85202049; PubMed=2704625;
RA Amuro N., Ooki K., Ito A., Goto Y., Okazaki T.;
RT "Nucleotide sequence of rat liver glutamate dehydrogenase cDNA.";
RL Nucleic Acids Res. 17:2356-2357(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WIS705048; PubMed=2704624;
RA Medina J., Moerer P., Leners W.H.;
RT "Nucleotide sequence of rat liver glutamate dehydrogenase cDNA.";
RL Nucleic Acids Res. 17:2355-2356(1989).
RN [3]
RP SEQUENCE OF 1-30 FROM N.A.
RC STRAIN=WISTAR; TISSUE=Liver;

Query Match 32.9%; Score 48; DB 1; Length 558;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 13; Conservative 50; Mismatches 7; Indels 4; Gaps 2;
OY 4 GLYQGVPRAPETGTEARRHYD 25
DB 31 GWARGOPSAVPOGLTPVARRHYSEG 56
RESULT 3
ID INVB_ZYMO STANDARD; PRT; 413 AA.
AC Q60115; Q60117; Q60125;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 01-DEC-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR SUCRASE (EC 3.2.1.26) (BETA-FRUCTOFURANOSIDASE)
DE (INVERTASE) (SACCHARASE) (PROTEIN B46).
GN SACC OR INVB OR SUCE3.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.
RN [1]
RP TaxID=542;
RC STRAIN=ATCC 10988 / ZM1;
RX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=94368848; PubMed=8086457;
RA Song K.B., Lee S.K., Joo H.K., acid sequences of an extracellular
RT "Nucleotide and deduced amino acid sequences of an extracellular
RL Nucleic Acids Res. 121:163-166(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 13756;
RX MEDLINE=95218269; PubMed=7766026;

Query Match 32.9%; Score 48; DB 1; Length 558;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 13; Conservative 50; Mismatches 7; Indels 4; Gaps 2;
OY 4 GLYQGVPRAPETGTEARRHYD 25
DB 31 GWARGOPSAVPOGLTPVARRHYSEG 56
RESULT 3
ID INVB_ZYMO STANDARD; PRT; 413 AA.
AC Q60115; Q60117; Q60125;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 01-DEC-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR SUCRASE (EC 3.2.1.26) (BETA-FRUCTOFURANOSIDASE)
DE (INVERTASE) (SACCHARASE) (PROTEIN B46).
GN SACC OR INVB OR SUCE3.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.
RN [1]
RP TaxID=542;
RC STRAIN=ATCC 10988 / ZM1;
RX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=94368848; PubMed=8086457;
RA Song K.B., Lee S.K., Joo H.K., acid sequences of an extracellular
RT "Nucleotide and deduced amino acid sequences of an extracellular
RL Nucleic Acids Res. 121:163-166(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 13756;
RX MEDLINE=95218269; PubMed=7766026;

Query Match 32.9%; Score 48; DB 1; Length 558;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 13; Conservative 50; Mismatches 7; Indels 4; Gaps 2;
OY 4 GLYQGVPRAPETGTEARRHYD 25
DB 31 GWARGOPSAVPOGLTPVARRHYSEG 56
RESULT 3
ID INVB_ZYMO STANDARD; PRT; 413 AA.
AC Q60115; Q60117; Q60125;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 01-DEC-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR SUCRASE (EC 3.2.1.26) (BETA-FRUCTOFURANOSIDASE)
DE (INVERTASE) (SACCHARASE) (PROTEIN B46).
GN SACC OR INVB OR SUCE3.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.
RN [1]
RP TaxID=542;
RC STRAIN=ATCC 10988 / ZM1;
RX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=94368848; PubMed=8086457;
RA Song K.B., Lee S.K., Joo H.K., acid sequences of an extracellular
RT "Nucleotide and deduced amino acid sequences of an extracellular
RL Nucleic Acids Res. 121:163-166(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 13756;
RX MEDLINE=95218269; PubMed=7766026;

Query Match 32.9%; Score 48; DB 1; Length 558;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 13; Conservative 50; Mismatches 7; Indels 4; Gaps 2;
OY 4 GLYQGVPRAPETGTEARRHYD 25
DB 31 GWARGOPSAVPOGLTPVARRHYSEG 56
RESULT 3
ID INVB_ZYMO STANDARD; PRT; 413 AA.
AC Q60115; Q60117; Q60125;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 01-DEC-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR SUCRASE (EC 3.2.1.26) (BETA-FRUCTOFURANOSIDASE)
DE (INVERTASE) (SACCHARASE) (PROTEIN B46).
GN SACC OR INVB OR SUCE3.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.
RN [1]
RP TaxID=542;
RC STRAIN=ATCC 10988 / ZM1;
RX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=94368848; PubMed=8086457;
RA Song K.B., Lee S.K., Joo H.K., acid sequences of an extracellular
RT "Nucleotide and deduced amino acid sequences of an extracellular
RL Nucleic Acids Res. 121:163-166(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 13756;
RX MEDLINE=95218269; PubMed=7766026;

Query Match 32.9%; Score 48; DB 1; Length 558;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 13; Conservative 50; Mismatches 7; Indels 4; Gaps 2;
OY 4 GLYQGVPRAPETGTEARRHYD 25
DB 31 GWARGOPSAVPOGLTPVARRHYSEG 56
RESULT 3
ID INVB_ZYMO STANDARD; PRT; 413 AA.
AC Q60115; Q60117; Q60125;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 01-DEC-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR SUCRASE (EC 3.2.1.26) (BETA-FRUCTOFURANOSIDASE)
DE (INVERTASE) (SACCHARASE) (PROTEIN B46).
GN SACC OR INVB OR SUCE3.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.
RN [1]
RP TaxID=542;
RC STRAIN=ATCC 10988 / ZM1;
RX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=94368848; PubMed=8086457;
RA Song K.B., Lee S.K., Joo H.K., acid sequences of an extracellular
RT "Nucleotide and deduced amino acid sequences of an extracellular
RL Nucleic Acids Res. 121:163-166(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 13756;
RX MEDLINE=95218269; PubMed=7766026;

Query Match 32.9%; Score 48; DB 1; Length 558;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 13; Conservative 50; Mismatches 7; Indels 4; Gaps 2;
OY 4 GLYQGVPRAPETGTEARRHYD 25
DB 31 GWARGOPSAVPOGLTPVARRHYSEG 56
RESULT 3
ID INVB_ZYMO STANDARD; PRT; 413 AA.
AC Q60115; Q60117; Q60125;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 01-DEC-2000 (Rel. 40, Last annotation update)
DE EXTRACELLULAR SUCRASE (EC 3.2.1.26) (BETA-FRUCTOFURANOSIDASE)
DE (INVERTASE) (SACCHARASE) (PROTEIN B46).
GN SACC OR INVB OR SUCE3.
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.
RN [1]
RP TaxID=542;
RC STRAIN=ATCC 10988 / ZM1;
RX SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE=94368848; PubMed=8086457;
RA Song K.B., Lee S.K., Joo H.K., acid sequences of an extracellular
RT "Nucleotide and deduced amino acid sequences of an extracellular
RL Nucleic Acids Res. 121:163-166(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 13756;
RX MEDLINE=95218269; PubMed=7766026;

Query Match 32.9%; Score 48; DB 1; Length 558;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 13; Conservative 50; Mismatches 7; Indels 4; Gaps 2;
OY 4 GLYQGVPRAPETGTEARRHYD 25
DB 31 GWARGOPSAVPOGLTPVARRHYSEG 56
RESULT 3
ID INVB_ZYMO STANDARD; PRT; 413 AA.
AC Q

RX MEDLINE-93162045; PubMed-1286669;
 RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,
 RA Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,
 RA Appel R.D., Hughes G.J.;
 RT "Human liver protein map: a reference database established by
 RT microsequencing and gel comparison.";
 RL Electrophoresis 13:992-1001(1992).
 RN [10]
 RP SUBSTRATE-BINDING SITE.
 RX MEDLINE-74125870; PubMed-4856315;
 RA Rasched I., Joernvall H., Sund H.;
 RT "Studies of glutamate dehydrogenase. Identification of an amino group
 RT involved in the substrate binding.";
 RL Eur. J. Biochem. 41:603-606(1974).
 RN [11]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-99268411; PubMed-10338089;
 RA Meissner T., Beinbrech B., Mayatepek E.;
 RT "Congenital hyperinsulinism: molecular basis of a heterogeneous
 RT disease.";
 RL Hum. Mutat. 13:351-361(1999).
 RN [12]
 RP VARIANTS HHS LEU-498; SER-499; ASP-499; PRO-501 AND TYR-507.
 RX MEDLINE-98223108; PubMed-9571255;
 RA Stanley C.A., Lieu Y.K., Hsu B.Y.L., Burlina A.B., Greenberg C.R.,
 RA Hopwood N.J., Perlman K., Rich B.H., Zammarchi E., Poncz M.;
 RT "Hyperinsulinism and hyperammonemia in infants with regulatory
 RT mutations of the glutamate dehydrogenase gene.";
 RL New Engl. J. Med. 338:1352-1357(1998).
 RN [13]
 RP VARIANTS HHS LYS-318 AND ALA-349.
 RX MEDLINE-20102775; PubMed-10636977;
 RA Miki Y., Taki T., Ohura T., Kato H., Yanagisawa M., Hayashi Y.;
 RT "Novel missense mutations in the glutamate dehydrogenase gene in the
 RT congenital hyperinsulinism-hyperammonemia syndrome.";
 RL J. Pediatr. 136:69-72(2000).
 CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE + H(2)O + NAD(P)(+) =
 CC 2-OXOGLUTARATE + NH(3) + NAD(P)H.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
 CC -1- DISEASE: OVERACTIVITY OF GLUD1 IS ASSOCIATED WITH THE
 CC HYPERINSULINISM-HYPERAMMONEMIA SYNDROME (HHS). ELEVATED OXIDATION
 CC RATE OF GLUTAMATE TO ALPHA-KETOGLUTARATE STIMULATES INSULIN
 CC SECRETION IN THE PANCREATIC BETA CELLS, WHILE THEY IMPAIR
 CC DETOXIFICATION OF AMMONIUM IN THE LIVER.
 CC -1- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X07674; CAA30521.1; -
 DR EMBL; J03248; AAA52523.1; -
 DR EMBL; M20867; AAA52526.1; -
 DR EMBL; X07769; CAA30598.1; -
 DR EMBL; X66300; CAA46994.2; -
 DR EMBL; X66301; CAA46994.2; JOINED.
 DR EMBL; X66302; CAA46994.2; JOINED.
 DR EMBL; X66303; CAA46994.2; JOINED.
 DR EMBL; X66304; CAA46994.2; JOINED.
 DR EMBL; X66305; CAA46994.2; JOINED.
 DR EMBL; X66306; CAA46994.2; JOINED.
 DR EMBL; X66307; CAA46994.2; JOINED.
 DR EMBL; X66308; CAA46994.2; JOINED.
 DR EMBL; X66309; CAA46994.2; JOINED.
 DR EMBL; X66311; CAA46994.2; JOINED.
 DR EMBL; X66312; CAA46994.2; JOINED.
 DR EMBL; M37154; AAA52525.1; -
 DR EMBL; X67491; CAA47830.1; -

DR PIR; A28208; DEHUE.
 DR PIR; S00958; S00958.
 DR HSP; P80319; ICTM.
 DR SWISS-2DPAGE; P00367; HUMAN.
 DR MIM; 138130; -
 DR InterPro; IPR001625; -
 DR Pfam; PF00208; GLFV_dehydrog; 1.
 DR PRINTS; PR00082; GLFDHGRNASE.
 DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
 KW Oxidoreductase; NADP; Mitochondrion; Transit peptide; Polymorphism;
 KW Disease mutation; Multigene family.
 FT TRANSIT 1 53 MITOCHONDRION.
 FT CHAIN 54 558 GLUTAMATE DEHYDROGENASE 1.
 FT BINDING 84 84 SUBSTRATE.
 FT ACT_SITE 183 183
 FT VARIANT 318 318
 FT VARIANT 349 349
 FT VARIANT 380 380
 FT VARIANT 498 498
 FT VARIANT 499 499
 FT VARIANT 499 499
 FT VARIANT 501 501
 FT VARIANT 507 507
 FT SEQUENCE 558 AA; 61397 MW; A7319A840F57FBB2 CRC64;
 SQ
 Query Match 32.2%; Score 47; DB 1; Length 558;
 Best Local Similarity 44.4%; Pred. No. 31;
 Matches 12; Conservative 1; Mismatches 10; Indels 4; Gaps 1;
 QY 4 GLYQGVPRAP-----GTEARRHYDEGV 26
 DB 31 GWARGQPAAPQPGALAAARRHYSEAV 57
 RESULT 6
 ILVE_ECOLI STANDARD; PRT; 308 AA.
 ID ILVE_ECOLI
 AC P00510; Q47299;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE BRANCHED-CHAIN AMINO ACID AMINOTRANSFERASE (EC 2.6.1.42) (TRANSAMINASE
 DE B) (BCAT).
 GN ILVE.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE-85289113; PubMed-3897211;
 RA Kuramitsu S., Ogawa T., Ogawa H., Kagamiyama H.;
 RT "Branched-chain amino acid aminotransferase of Escherichia coli:
 RT nucleotide sequence of the ilvE gene and the deduced amino acid
 RT sequence.";
 RL J. Biochem. 97:993-999(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE-87174741; PubMed-3550695;
 RA Lawther R.P., Wek R.C., Lopes J.M., Pereira R., Taillon B.E.,
 RA Hatfield G.W.;
 RT "The complete nucleotide sequence of the ilvGMDA operon of

Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
KW Pyridoxal phosphate; 3D-structure.
FT INIT_MET 0 1 THR MODIFIED TO GLY.
FT BINDING 159 159 PYRIDOXAL PHOSPHATE.
FT CONFLICT 150 150 A -> R (IN REF. 3).
FT SEQUENCE 308 AA; 33962 MW; 28404978DB79659 CRC64;
SQ

Query Match 31.8%; Score 46.5; DB 1; Length 308;
Best Local Similarity 31.4%; Pred. NO. 20;
Matches 16; Conservative 3; Mismatches 5; Indels 27; Gaps 3;
QY 3 EGIYQGV-----RAEP-----GTEARRH-YDEGV 26
DB 133 EALEGGIDAMVSSWNRRAAPNTIPTAAKAGNYLLSGSEARRHGYOEGI 183

RESULT 7
FAN_HUMAN ID FAN_HUMAN STANDARD; PRT; 917 AA.
AC Q92636; 1998 (Rel. 36, Created)
AC Q92636; 1998 (Rel. 36, Last sequence update)
DT 13-JUL-1998 (Rel. 30, Last annotation update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PROTEIN FAN (FACTOR ASSOCIATED WITH N-SMASE ACTIVATION) (FACTOR
DE ASSOCIATED WITH NEUTRAL-SPHINGOMYELINASE ACTIVATION).
DE NSMAF OR FAN.
OS Homo sapiens (Human).
OS Eukaryota; Eukaryota; Chordata; Granata; Vertebrata; Euteleostomi;
OC Eukaryota; Eukaryota; Chordata; Granata; Vertebrata; Euteleostomi;
OC NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RS TISSUE=Muscle;
RX MEDLINE=96404447; Pubmed=8808629;
RA Adam-Klages S., Adam D., Wiedmann K., Struve S., Kolanus W.,
RA Schneider-Mergener J., Krenn M., Krenn M., Krenn M., Krenn M.,
RA neutral sphingomyelinase.;
RL Cell 86:937-947(1996).
CC -!- FUNCTION: COUPLES THE P55 TNF-RECEPTOR (TNF-R55 / TNFR1) TO THE N-
CC NEUTRAL SPHINGOMYELINASE (N-SMASE). SPECIFICALLY BINDS TO THE N-
CC SMASE ACTIVATION DOMAIN OF TNF-R55. MAY REGULATE CERAMIDE
CC PRODUCTION BY N-SMASE-INDUCED TISSUES.
CC -!- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: CONTAINS 1 BEACH DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions, as long as the statement in no way
CC infringes on the intellectual property rights of the EMBL and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X96586; CAA65405.1; -
DR MIN; 603043;
DR INTERPRO; IPR001680; -
DR PRINTS; PR00320; GPOTELNBRPT.
DR PROSITE; PS00197; BEACH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
DR PROSITE; PS50082; WD_REPEATS_2; 4.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR Repeat; WD repeat.
KW DOMAIN; 209 575
FT BEACH.
FT REPEAT 670 700
FT REPEAT 712 740
FT REPEAT 761 791
FT REPEAT 803 833

```

FT REPEAT      884   914   WD 6.
FT DOMAIN      23   28   POLY-LEU.
SQ SEQUENCE    917 AA; 104317 MW; D8498B5FB6D8E139 CRC64;

Query Match      31.8%; Score 46.5; DB 1; Length 917;
Best Local Similarity 47.4%; Pred. NO. 59;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYXGVPRAPGTEARRHYD 23
DB 738 VWSGVPAEMPGRK-RHED 755

RESULT 8
SURE_HAEIN STANDARD; PRT; 249 AA.
AC P45681;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE STATIONARY-PHASE SURVIVAL PROTEIN SURE HOMOLOG.
GN SURE OR H10702.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -!- FUNCTION: NOT KNOWN; COULD BE A PHOSPHATASE.
CC -!- SIMILARITY: BELONGS TO THE SURE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U32753; AAC22361.1; -
DR TIGR: H10702; -
DR InterPro: IPR002828; -
DR Pfam: PF01975; Sure; 1.
SQ SEQUENCE 249 AA; 27340 MW; D10F280C95266757 CRC64;

Query Match      31.5%; Score 46; DB 1; Length 249;
Best Local Similarity 32.0%; Pred. NO. 19;
Matches 8; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 3 ECLXGVPRAPGTEARRHYDEGVR 27
DB 111 EGRHGLPAIVSLDGRQHYETAAR 135

RESULT 9
LAC4_THACU STANDARD; PRT; 531 AA.
ID LAC4_THACU

```

```

Q02081;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
15-JUL-1999 (Rel. 38, Last annotation update)
DE LACASE 4 PRECURSOR (EC 1.10.3.2) (BENZENEDIOL-OXYGEN OXIDOREDUCTASE)
DE (URISHIOL OXIDASE) (DIPHENOL OXIDASE).
GN LCC4.
OS Thanatephorus cucumeris (Black scurf of potato) (Rhizoctonia solani).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Ceratobasidiales;
OC Ceratobasidiaceae; mitosporic Ceratobasidiaceae; Rhizoctonia.
OX NCBI_TaxID=107832;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=RS22;
RX MEDLINE=96171523; PubMed=8598061;
RA Wahleithner J.A., Xu F., Brown K.M., Brown S.H., Golightly E.J.,
RA Halkier T., Kauppinen S., Pederson A., Schneider P.;
RT "The identification and characterization of four laccases from the
RT plant pathogenic fungus Rhizoctonia solani."
RL Curr. Genet. 29:395-403(1996).
CC -!- FUNCTION: LIGNIN DEGRADATION AND DETOXIFICATION OF LIGNIN-DERIVED
CC PRODUCTS (PROBABLE). THIS ISOZYME SHOWS OPTIMAL ACTIVITY AT PH 7.
CC -!- CATALYTIC ACTIVITY: 4 BENZENEDIOL + O(2) = 4 BENZOSEMIQUINONE +
CC 2 H(2)O.
CC -!- COFACTOR: BINDS 4 CU-IONS PER MOLECULE. THREE DISTINCT CU
CC CENTERS KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE
CC 3 OR COUPLED BINUCLEAR (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- TISSUE SPECIFICITY: IN MYCELIA, AT A HIGHER LEVEL THAN LCC1, LCC2
CC AND LCC3.
CC -!- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -!- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL: 254277; CAA91042.1; -
InterPro: IPR001117; -
Pfam: PF00394; Cu-oxidase; 2.
PROSITE: PS00079; MULTICOPPER_OXIDASE1; FALSE NEG.
PROSITE: PS00080; MULTICOPPER_OXIDASE2; FALSE NEG.
OXidoreductase; Signal: Copper; Metal-binding; Glycoprotein; Repeat;
Lignin degradation; Multigene family; Polymorphism.
SIGNAL 1 19
CHAIN 20 531
FT DOMAIN 23 146 PLASTOCYANIN-LIKE 1.
FT DOMAIN 158 315 PLASTOCYANIN-LIKE 2.
FT DOMAIN 384 507 PLASTOCYANIN-LIKE 3.
FT METAL 83 83 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 85 85 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 128 128 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 130 130 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 427 427 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 430 430 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 432 432 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 479 479 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 480 480 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 481 481 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 484 484 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL ? ? COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 66 66 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 109 109 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 395 395 N-LINKED (GLCNAC. .) (POTENTIAL).

```

•

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RC SEQUENCE FROM N.A., AND VARIANTS FMEN1.
 RC TISSUE=Leukocyte;
 RX MEDLINE=97258940; PubMed=9103196;
 RA Chandrasekharappa S.C., Guru S.C., Manickam P., Olufemi S.-E.,
 RA Collins F.S., Emmert-Buck M.R., Debelenko L.V., Zhuang Z.,
 RA Lubensky I.A., Liotta L.A., Crabtree J.S., Wang Y., Roe B.A.,
 RA Weisemann J., Boquist M.A., Agarwal S.K., Kester M.B., Kim Y.S.,
 RA Heppner C., Dong Q., Spiegel A.M., Burns A.L., Marx S.J.;
 RT "Positional cloning of the gene for multiple endocrine neoplasia-type
 RT 1.";
 RL Hum. Mol. Genet. 6:1169-1175(1997).
 RN [2]
 RN VARIANTS FMEN1 AND SPORADIC MEN1.
 RX MEDLINE=97358593; PubMed=9215689;
 RA Agarwal S.K., Kester M.B., Debelenko L.V., Heppner C.,
 RA Emmert-Buck M.R., Skarulis M.C., Doppman J.L., Kim Y.S.,
 RA Lubensky I.A., Zhuang Z., Green J.S., Guru S.C., Manickam P.,
 RA Olufemi S.E., Liotta L.A., Chandrasekharappa S.C., Collins F.S.,
 RA Spiegel A.M., Burns A.L., Marx S.J.;
 RT "Germline mutations of the MEN1 gene in familial multiple endocrine
 RT neoplasia type 1 and related states.";
 RL Hum. Mol. Genet. 6:1169-1175(1997).
 RN [3]
 RN VARIANT PARATHYROID ADENOMA LYS-26.
 RX MEDLINE=97385243; PubMed=9241276;
 RA Heppner C., Kester M.B., Agarwal S.K., Debelenko L.V.,
 RA Emmert-Buck M.R., Guru S.C., Manickam P., Olufemi S.-E.,
 RA Skarulis M.C., Doppman J.L., Alexander R.H., Kim Y.S., Saggar S.K.,
 RA Lubensky I.A., Zhuang Z., Liotta L.A., Chandrasekharappa S.C.,
 RA Collins F.S., Spiegel A.M., Burns A.L., Marx S.J.;
 RT "Somatic mutation of the MEN1 gene in parathyroid tumours.";
 RL Nat. Genet. 16:375-378(1997).
 RN [4]
 RN VARIANTS FMEN1 D-42; P-165; D-169; S-188 AND E-289.
 RX MEDLINE=98130524; PubMed=9463336;
 RA Bassett J.H.D., Forbes S.A., Pannett A.A.J., Lloyd S.E.,
 RA Christie P.T., Wooding C., Harding B., Besser G.M., Edwards C.R.,
 RA Monson J.P., Sampson J., Wass J.A.H., Wheeler M.H., Thakker R.V.;
 RT "Characterization of mutations in patients with multiple endocrine
 RT neoplasia type 1.";
 RL Am. J. Hum. Genet. 62:232-244(1998).
 RN [5]
 RN VARIANTS FMEN1.
 RX MEDLINE=98349969; PubMed=9683585;
 RA Giraud S., Zhang C.X., Serova-Sinilnikova O., Wautot V., Salandre J.,
 RA Buliss N., Waterlot C., Batters C., Porchet N., Aubert J.-P., Emy P.,
 RA Cadot G., Delmer B., Chabre O., Niccoli P., Lepret F., Duron F.,
 RA Emperauger B., Cougard P., Goudet P., Sarfati E., Riou J.-P.,
 RA Guichard S., Rodier M., Meyrier A., Caron P., Vantghem M.-C.,
 RA Assayag P., Peix J.-L., Pugeat M., Rohmer V., Vallotton M., Lenoir G.,
 RA Gaudray P., Murat A., Calender A.;
 RT "Germline mutation analysis in patients with multiple endocrine
 RT neoplasia type 1 and related disorders.";
 RL Am. J. Hum. Genet. 63:455-467(1998).
 RN [6]
 RN VARIANT FIHP LYS-260.
 RX MEDLINE=99011276; PubMed=9792884;
 RA Teh B.T., Espaa C.T., Houlston R., Grandell U., Farnebo F.,
 RA Nordenskjöld M., Pearce C.J., Carmichael D., Larsson C., Harris P.E.;
 RT "A family with isolated hyperparathyroidism segregating a missense
 RT MEN1 mutation and showing loss of the wild-type alleles in the
 RT parathyroid tumors.";
 RL Am. J. Hum. Genet. 63:1544-1549(1998).
 RN [7]
 RN VARIANT FIHP GLU-189.
 RX MEDLINE=99057176; PubMed=9843042;
 RA Fujimori M., Shirahama S., Anano J., Hashizume K., Ito K.,
 RA Shingu K., Kobayashi S., Anano J., Fukushima Y.;
 RT "Novel V184E MEN1 germline mutation in a Japanese kindred with

RT familial hyperparathyroidism.";
 RL Am. J. Med. Genet. 80:221-222(1998).
 RN [8]
 RN VARIANTS FMEN1.
 RX MEDLINE=98334342; PubMed=9671267;
 RA Agarwal S.K., Debelenko L.V., Kester M.B., Guru S.C., Manickam P.,
 RA Olufemi S.-E., Skarulis M.C., Heppner C., Crabtree J.S.,
 RA Lubensky I.A., Zhuang Z., Kim Y.S., Chandrasekharappa S.C.,
 RA Collins F.S., Liotta L.A., Spiegel A.M., Burns A.L., Emmert-Buck M.R.,
 RA Marx S.J.;
 RT "Analysis of recurrent germline mutations in the MEN1 gene encountered
 RT in apparently unrelated families.";
 RL Hum. Mutat. 12:75-82(1998).
 RN [9]
 RN VARIANT FMEN1 ILE-135 AND LYS-364.
 RX MEDLINE=98410971; PubMed=9740255;
 RA Boeni R., Vortmeyer A.O., Pack S., Park W.-S., Burg G., Hofbauer G.,
 RA Darling T., Liotta L., Zhuang Z.;
 RT "Somatic mutations of the MEN1 tumor suppressor gene detected in
 RT sporadic angiofibromas.";
 RL J. Invest. Dermatol. 111:539-540(1998).
 RN [10]
 RN VARIANTS FMEN1 LYS-119 DEL AND GLN-171--LEU-173 DEL.
 RX MEDLINE=98419173; PubMed=9747036;
 RA Sakurai A., Shirahama S., Fujimori M., Katai M., Itakura Y.,
 RA Kobayashi S., Anano J., Fukushima Y., Hashizume K.;
 RT "Novel MEN1 gene mutations in familial multiple endocrine neoplasia
 RT type 1.";
 RL J. Hum. Genet. 43:199-201(1998).
 RN [11]
 RN VARIANT FMEN1 GLY-45.
 RX MEDLINE=99048878; PubMed=9832038;
 RA Sato M., Matsubara S., Miyauchi A., Ohye H., Imachi H., Murao K.,
 RA Takahara J.;
 RT "Identification of five novel germline mutations of the MEN1 gene in
 RT Japanese multiple endocrine neoplasia type 1 (MEN1) families.";
 RL J. Med. Genet. 35:915-919(1998).
 RN [12]
 RN VARIANTS FMEN1 TRP-39; TYR-177; ASP-184; PRO-269 AND PRO-272.
 RX MEDLINE=99103464; PubMed=9888389;
 RA Poncin J., Abs R., Veikeniers B., Bonduelle M., Abramowicz M.,
 RA Legros J.-J., Verloes A., Meurisse M., van Gaal L., Verellen C.,
 RA Koulischer L., Beckers A.;
 RT "Mutation analysis of the MEN1 gene in Belgian patients with multiple
 RT endocrine neoplasia type 1 and related diseases.";
 RL Hum. Mutat. 13:54-60(1999).
 RN [13]
 RN VARIANTS MEN1 ASP-161 AND ARG-246.
 RX MEDLINE=99188881; PubMed=10090472;
 RA Mutch M.G., Dilley W.G., Sanjurjo F., Debenedetti M.K., Doherty G.M.,
 RA Wells S.A., Jr., Goodfellow P.J., Laimore T.C.;
 RT "Germline mutations in the multiple endocrine neoplasia type 1 gene:
 RT evidence for frequent splicing defects.";
 RL Hum. Mutat. 13:175-185(1999).
 CC -1- FUNCTION: NOT KNOWN.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.
 CC -1- DISEASE: DEFECTS IN MEN1 IS THE CAUSE OF FAMILIAL MULTIPLE
 CC ENDOCRINE NEOPLASIA TYPE 1 (FMEN1); WERNER SYNDROME, AN AUTOSOMAL
 CC DOMINANT DISORDER CHARACTERIZED BY TUMORS OF THE PARATHYROID
 CC GLANDS, GASTRO-INTESTINAL ENDOCRINE TISSUE, THE ANTERIOR PITUITARY
 CC AND OTHER TISSUES. CUTANEOUS LESIONS AND NERVOUS-TISSUE TUMORS
 CC CAN EXIST. PROGNOSIS IN FMEN1 PATIENTS IS RELATED TO HORMONAL
 CC HYPERSECRETION BY TUMORS, SUCH AS HYPERGASTRINEMIA CAUSING SEVERE
 CC PEPTIC ULCER DISEASE (ZOLLINGER-ELISON SYNDROME, ZES), PRIMARY
 CC HYPERPARATHYROIDISM, AND ACUTE FORMS OF HYPERINSULINEMIA.
 CC -1- DISEASE: DEFECTS IN MEN1 IS THE CAUSE OF FAMILIAL ISOLATED
 CC HYPERPARATHYROIDISM (FIHP OR HRPT1). FIHP IS AN AUTOSOMAL DOMINANT
 CC DISORDER CHARACTERIZED BY HYPERCALCEMIA, ELEVATED PARATHYROID
 CC HORMONE (PTH) LEVELS, AND UNIGLANDULAR OR MULTIGLANDULAR
 CC PARATHYROID TUMORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC HSV-2 UL54, EBV-1 5, VZV 4, EBV BMLF1, HCMV UL69, AND HVS-1 57.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X17403; CAA35384.1; -
 DR PIR: S09832; Q0BEA7.
 KW Transcription regulation.
 SQ SEQUENCE 744 AA; 82678 MW; D0566B2009EB5828 CRC64;

Query Match 31.5%; Score 46; DB 1; Length 744;
 Best Local Similarity 43.8%; Pred. No. 56;
 Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
 QY 10 PRAEPTGTEARRHYDEG 25
 | | | | : : : | | |
 Db 598 PRAQPPSQQHYSEG 613

RESULT 15
 ILVE_SALTY STANDARD; PRT; 308 AA.
 AC P15168;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE BRANCHED-CHAIN AMINO ACID AMINOTRANSFERASE (EC 2.6.1.42) (TRANSAMINASE
 DE B) (BCAT).
 GN ILVE.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=89352621; PubMed=2669973;
 RA Feild M.J., Nguyen D.C., Armstrong F.B.;
 RT "Amino acid sequence of Salmonella typhimurium branched-chain amino
 RT acid aminotransferase".
 RL Biochemistry 28:5306-5310(1989).
 RN [2]

SEQUENCE OF 1-5.
 RA Randall R.R., Wallis M.H., Young G.J., Armstrong F.B.;
 RT "N-terminal sequence of branched-chain amino acid aminotransferase";
 RL Fed. Proc. 38:325-325(1979).
 CC -1- CATALYTIC ACTIVITY: L-LEUCINE + 2-OXOGLUTARATE -> 4-METHYL-2-
 CC OXOPENTANOATE + L-GLUTAMATE (ALSO ACTS ON L-ISOLEUCINE AND
 CC L-VALINE).
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
 CC -1- PATHWAY: VALINE AND ISOLEUCINE BIOSYNTHESIS.
 CC -1- SUBUNIT: HOMOHXAMER.
 CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES.
 DR PIR: A05077; A05077.
 DR PIR: A34082; A34082.
 DR HSP: P00510; 1A3G.
 DR StyGene; SG10181; ilve.
 DR InterPro; IPR001544; -.
 DR Pfam; PF01063; aminotran_4; 1.
 DR PROSITE; PS00770; AA_TRANSF_CLASS_4; 1.
 KW Transferase; Aminotransferase; Branched-chain amino acid biosynthesis;
 KW Pyridoxal phosphate.
 FT BINDING 159 159 PYRIDOXAL PHOSPHATE.
 SQ SEQUENCE 308 AA; 33921 MW; 8F826434F9588AAD CRC64;

Query Match 31.2%; Score 45.5; DB 1; Length 308;

Best Local Similarity 31.4%; Pred. No. 27;
 Matches 16; Conservative 3; Mismatches 5; Indels 27; Gaps 3;
 QY 3 EGLYQGVY-----RAEP-----GTEARRH-YDEGV 26
 | | | | : : : | | | | | | | | | |
 Db 133 EALDQGDAMVSSWNRAAPNTIPTAAKAGNYLSSLLVGSARRHGQEG 183
 RESULT 16
 EST5_DROMO STANDARD; PRT; 38 AA.
 AC P10095;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE ESTERASE 5 (EC 3.1.1.1) (FRAGMENT).
 GN EST5 OR EST-5.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7230;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=87100025; PubMed=3800957;
 RA Pen J., van Beeumen J., Beintema J.J.;
 RT "Structural comparison of two esterases from Drosophila melanogaster
 RT isolated by immunoaffinity chromatography".
 RL Biochem. J. 238:691-699(1986).
 CC -1- CATALYTIC ACTIVITY: A CARBOXYLIC ESTER + H(2)O -> AN ALCOHOL
 CC + A CARBOXYLIC ANION.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 DR PIR: B29502; B29502.
 DR FlyBase; FBgn0012569; Dmoj\Est-5.
 DR InterPro; IPR002018; -.
 DR Pfam; PF00135; Coesterase; 1.
 DR PROSITE; PS00122; CARBOXYLESTERASE_B_1; PARTIAL.
 DR PROSITE; PS00941; CARBOXYLESTERASE_B_2; PARTIAL.
 KW Hydrolase; Serine esterase.
 FT NON_TER 38 38
 SQ SEQUENCE 38 AA; 4008 MW; 61A3C46CF084A465 CRC64;

Query Match 30.8%; Score 45; DB 1; Length 38;
 Best Local Similarity 64.3%; Pred. No. 4;
 Matches 9; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 3 EGLY--QGVRAEP 14
 | | | : | | | | |
 Db 22 EGYEAEIGIPRAEP 35
 RESULT 17
 FASA_RAT STANDARD; PRT; 324 AA.
 ID FASA_RAT
 AC O63199;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE FASL RECEPTOR PRECURSOR (APOPTOSIS-MEDIATING SURFACE ANTIGEN FAS)
 DE (APO-1 ANTIGEN) (CD95).
 GN TNFRSF6 OR PT1 OR FAS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
 RX MEDLINE=94128114; PubMed=7507668;
 RA Kimura K., Yamamoto M., Wakatsuki T.;
 RT "A variant mRNA species encoding a truncated form of Fas antigen in
 RT the rat liver".

FT	DOMAIN	2040	2098	SUSHI.
FT	DOMAIN	48	140	G1-A.
FT	DOMAIN	152	247	G1-B.
FT	DOMAIN	253	349	G1-B'.
FT	DOMAIN	486	580	G2-B.
FT	DOMAIN	587	682	G2-B'.
FT	DOMAIN	685	798	KS.
FT	DOMAIN	801	1226	CS-1.
FT	DOMAIN	1227	1909	CS-2.
FT	DOMAIN	1910	2124	G3.
FT	DISULFID	51	133	BY SIMILARITY.
FT	DISULFID	175	246	BY SIMILARITY.
FT	DISULFID	199	220	BY SIMILARITY.
FT	DISULFID	273	348	BY SIMILARITY.
FT	DISULFID	297	318	BY SIMILARITY.
FT	DISULFID	509	580	BY SIMILARITY.
FT	DISULFID	533	554	BY SIMILARITY.
FT	DISULFID	607	682	BY SIMILARITY.
FT	DISULFID	631	652	BY SIMILARITY.
FT	DISULFID	1914	1925	BY SIMILARITY.
FT	DISULFID	1942	2034	BY SIMILARITY.
FT	DISULFID	2010	2026	BY SIMILARITY.
FT	DISULFID	2041	2084	BY SIMILARITY.
FT	DISULFID	2070	2097	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	387	387	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	611	611	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	667	667	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1842	1842	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	SEQUENCE	2124 AA; 221117 MW;	E30BBB61593A34B1 CRC64;	

Query Match 30.8%; Score 45; DB 1; Length 2124;
 Best Local Similarity 43.5%; Pred. No. 2.2e+02;
 Matches 10; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 4 GLYQGVPRAEPTGEARRHYDEGV 26
 | - | - | - | - | - | - |
 DB 1100 GYVSGIPSDGDTSTSTSGVEGV 1122

RESULT	20
KAD_PARDE	
ID	KAD_PARDE STANDARD; PRT; 217 AA.
AC	P10772; P77922;
DT	01-JUL-1989 (Rel. 11, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	15-JUL-1998 (Rel. 36, Last annotation update)
DE	ADENYLATE KINASE (EC 2.7.4.3) (ATP-AMP TRANSPHOSPHORYLASE).
GN	ADK.
OS	Paracoccus denitrificans.
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OX	Paracoccus.
NCBI_TaxID=266;	
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=71.llt;
RC	MEDLINE=99077027; PubMed=9862211;
RA	Perrier V., Burlacu-Miron S., Boussac A., Meier A., Gilles A.M.;
RX	"Metal chelating properties of adenylate kinase from Paracoccus
RT	denitrificans";
RT	adenylate kinases.";
RL	Protein Eng. 11:917-923(1998).
RP	[2]
RN	SEQUENCE.
RA	MEDLINE=89153098; PubMed=2537726;
RX	Spurgin P., Tomasselli A.G., Schiltz E.;
RT	"The amino acid sequence of adenylate kinase from Paracoccus
RT	denitrificans and its relationship to mitochondrial and microbial
CC	-!- FUNCTION: THIS SMALL UBQUITOUS ENZYME IS ESSENTIAL FOR

CC MAINTENANCE AND CELL GROWTH. ADP + AMP = ADP + ADP.
CC -!- CATALYTIC ACTIVITY: ATP + AMP = ADP + ADP.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE ADENYLATE KINASE FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U64203; AB06328.1; -
CC DR PIR; S04826; S04820.
CC DR HSP; P05082; 2ECK.
CC DR InterPro; IPR000850; -
CC DR Pfam; PF00406; adenylatekinase; 1.
CC DR PRINTS; PR00094; ADENYLATE_KINASE; 1.
CC DR PROSITE; PS00113; ADENYLATE_KINASE; 1.
CC DR TRANSFERASE; Kinase; 6
CC KW BINDING; ATP-binding.
CC FT NPI_BIND 8 16 ATP (BY SIMILARITY).
FT DISULFID 149 152
FT CONFLICT 129 129 C -> H (IN REF. 2).
FT CONFLICT 137 137 H -> G (IN REF. 2).
FT CONFLICT 217 217 9A2C78AADS1A37C6 CRC64;
FT SEQUENCE 217 AA; 23680 MW; 9A2C78AADS1A37C6 CRC64;
CC
CC Query Match 30.1%; Score 44; DB 1; Length 217;
CC Best Local Similarity 58.8%; Pred. No. 31;
CC Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
CC
CC QY 8 GVPRAEPTARRHYDE 24
CC DB 8 GPGRGKGTQARRLIDE 24
CC
CC RESULT 21
CC LCB2_KLULA STANDARD; PRT; 562 AA.
CC AC P48241; 1996 (Rel. 33, Created)
CC DT 01-FEB-1996 (Rel. 33, Last sequence update)
CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC DE SERINE PALMITOYLTRANSFERASE 2 (EC 2.3.1.50) (LONG CHAIN BASE
CC DE BIOSYNTHESIS PROTEIN 2) (SPT 2).
CC GN LCB2.
CC OS Kluyveromyces fragilis (Yeast).
CC CC Eukaryota; Fungi; Ascomycota; Saccharomycetes;
CC CC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Kluyveromyces.
CC OX NCBI_TaxID=28985;
CC RN SEQUENCE FROM N.A.
CC RC STRAIN-JAG;
CC RX MEDLINE-97080528; PubMed-8921873; R.C.;
CC RA Nagiec M.M., Lester R.L., Dicksen R.L.;
CC RT Splicing of the Lcb2 gene: identification and characterization of
CC RT multiple cDNAs encoding the Lcb2 subunit of serine
CC RT palmitoyltransferase.";
CC RL Gene 177:237-241(1996).
CC CC -!- CATALYTIC ACTIVITY: PALMITOYL-COA + L-SERINE - COA + 3-DEHYDRO-
CC CC D-SPHINGAMINE + CO(2).
CC CC -!- COFACTOR: IDOPHOSPHATE.
CC CC -!- COMPONENT: FIRST STEP IN THE BIOSYNTHESIS OF THE LONG-CHAIN BASE
CC CC -!- SUBCELLULAR LOCATION: SPHINGOLIPIDS.
CC CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC CC -!- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC CC AMINOTRANSFERASES.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U15646; AAC49535.1; -
CC DR InterPro; IPR001917; -
CC DR Pfam; PF00222; aminotran_2; 1; CLASS 2; 1.
CC DR PROSITE; PS00599; AA-TRANSFER, Transmembrane; Pyridoxal phosphate.
CC KW TRANSFERASE; Cytitransf; 81
CC FT BINDING 365 365 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT SEQUENCE 562 AA; 63004 MW; 94249600A8BFD2F5 CRC64;
CC
CC Query Match 30.1%; Score 44; DB 1; Length 562;
CC Best Local Similarity 61.5%; Pred. No. 82;
CC Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
CC
CC QY 4 GLYCVPRPRAEPT 16
CC DB 186 GLYCVPRPRTIGT 198
CC
CC RESULT 22
CC PEC1_PIG STANDARD; PRT; 740 AA.
CC AC Q95242;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 05-JUL-2000 (Rel. 38, Last annotation update)
CC DE PLATELET ENDOTHELIAL CELL ADHESION MOLECULE PRECURSOR (PECAM-1)
CC DE (CD31 ANTIGEN).
CC GN PECAM1.
CC OS Sus scrofa (Pig).
CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
CC OX NCBI_TaxID=9823;
CC RN SEQUENCE FROM N.A.
CC RA Nasu K.;
CC RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases
CC CC -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE EXPRESSED ON
CC CC PLATELETS AND AT ENDOTHELIAL CELL-CELL JUNCTIONS.
CC CC -!- SUBCELLULAR LOCATION: TRANSMEMBRANE PROTEIN.
CC CC -!- SIMILARITY: BELONGS TO CLASS-6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X98505; CAA67129.1; -
CC DR InterPro; IPR003006; -
CC DR Pfam; PF00047; Ig; 4.
CC KW Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
CC FT SIGNAL.
CC FT CHAIN 1 27 BY SIMILARITY.
CC FT CHAIN 28 740 PLATELET ENDOTHELIAL CELL ADHESION
CC FT MOLECULE.
CC FT DOMAIN 28 602 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 603 621 POTENTIAL_C (POTENTIAL).
CC FT DOMAIN 622 740 C1-ESTERASE.
CC FT DOMAIN 740 740 IG-LIKE C2-TYPE DOMAIN.
CC FT DOMAIN 145 213 IG-LIKE C2-TYPE DOMAIN.
CC FT DOMAIN 249 311 IG-LIKE C2-TYPE DOMAIN.
CC FT DOMAIN 340 394 IG-LIKE C2-TYPE DOMAIN.
CC FT DOMAIN 425 484 IG-LIKE C2-TYPE DOMAIN.
CC

FT DOMAIN 517 580 IG-LIKE C2-TYPE DOMAIN.
 FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 357 357 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 436 436 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 456 456 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 740 AA; 82378 MW; F312DC62C4B4A217 CRC64;

Query Match 30.1%; Score 44; DB 1; Length 740;
 Best Local Similarity 63.6%; Pred. No. 1.1e+02;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 12 AEPGTEARRHY 22
 : : | | | | |
 Db 655 SDPNTANRHY 665

RESULT 23
 RPAL_HALHA STANDARD; PRT; 971 AA.
 ID RPAL_HALHA
 AC P15350;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE DNA-DIRECTED RNA POLYMERASE SUBUNIT A' (EC 2.7.7.6).
 GN RPOA1.
 OS Halobacterium halobium.
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=2241;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-R1;
 RX MEDLINE=89199633; PubMed=2495365;
 RA Leffers H., Gropp F., Lottspeich F., Zillig W., Garrett R.A.;
 RT "Sequence, organization, transcription and evolution of RNA
 RT polymerase subunit genes from the archaeobacterial extreme halophiles
 RT Halobacterium halobium and Halococcus morrhuae.";
 RL J. Mol. Biol. 206:1-17(1989).
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE +
 CC RNA(N).
 CC -!- COFACTOR: ZINC.
 CC -!- SUBUNIT: THIS RNA POLYMERASE IS COMPOSED OF 5 LARGE SUBUNITS:
 CC A', A'', B', B'' AND EPSILON, PRESENT IN STOICHIOMETRIC AMOUNTS, AND
 CC 3 SMALL ONES, E, F AND G, PRESENT IN SUB-STOICHIOMETRIC AMOUNTS.
 CC -!- SIMILARITY: THE COMBINED A'+A" SUBUNITS CORRESPOND TO THE A
 CC SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE
 CC EUKARYOTIC BETA' SUBUNIT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X57144; CAA0426.1; -;
 CC F01: S03574; S03574.
 CC InterPro: IPR000722; -;
 CC Pfam: PF00623; RNA_pol_A.1.
 CC Transferase; DNA-directed RNA polymerase; Transcription; Zinc.
 FT MOD_RES 1 1 BLOCKED.
 FT ZN_FING 62 105 C4-TYPE (POTENTIAL).

SQ SEQUENCE 971 AA; 108683 MW; FA91E23E8DEA9A84 CRC64;

Query Match 30.1%; Score 44; DB 1; Length 971;
 Best Local Similarity 53.3%; Pred. No. 1.4e+02;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 11 RAEPGTEARRHYDEG 25
 : : | | | | : | : |
 Db 471 RVEPGWEVQRHLIDG 485

RESULT 24
 VPG_BYDVN STANDARD; PRT; 153 AA.
 ID VPG_BYDVN
 AC P27579;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE PUTATIVE GENOME-LINKED PROTEIN PRECURSOR (VPG) (17 KDA PROTEIN).
 OS Barley yellow dwarf virus (isolate NY-RPV) (BYDV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Luteovirus.
 OX NCBI_TaxID=12039;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91108372; PubMed=2273382;
 RA Vincent J.R., Deng P.P., Lister R.M., Larkins B.A.;
 RT "Nucleotide sequences of coat protein genes for three isolates of
 RT barley yellow dwarf virus and their relationships to other luteovirus
 RT coat protein sequences.";
 RL J. Gen. Virol. 71:2791-2799(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92013950; PubMed=1840612;
 RA Larkins B.A., Lister R.M., Vincent J.R.;
 RT "Nucleotide sequence analysis and genomic organization of the NY-RPV
 RT isolate of barley yellow dwarf virus.";
 RL J. Gen. Virol. 72:2347-2355(1991).
 CC -!- FUNCTION: THIS PROTEIN IS POSSIBLY IS A VPG-PRECURSOR FROM WHICH,
 CC AT THE ONSET OF THE RNA SYNTHESIS, THE VPG MOLECULE IS RELEASED,
 CC AS IT HAS BEEN SUGGESTED FOR COMPEA MOSAIC VIRUS.
 CC -!- SIMILARITY: BELONGS TO THE LUTEOVIRUSES VPG PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D10205; BAA01056.1; -;
 CC EMBL: X17259; CAA35161.1; -;
 CC EMBL: L25299; AAA42870.1; -;
 CC InterPro: IPR001964; -;
 CC Pfam: PF01659; Luteo_Vpg; 1.
 CC PRINTS: PR00912; LVIRUSORF5.
 SQ SEQUENCE 153 AA; 17209 MW; 3839E5C502C51602 CRC64;

Query Match 29.8%; Score 43.5; DB 1; Length 153;
 Best Local Similarity 41.7%; Pred. No. 26;
 Matches 10; Conservative 3; Mismatches 4; Indels 7; Gaps 1;

QY 1 VGEGLYQ-----GVPRAPGTE 17
 : : | | | | : | : | | |
 Db 13 LGEGLLQERSQWLSLPTAQPGE 36

RESULT 25
 ASGX_PYRAB STANDARD; PRT; 305 AA.
 ID ASGX_PYRAB

```

CC CATALYSING BRANCH MIGRATION, HAS A DNA UNWINDING ACTIVITY
CC CHARACTERISTIC OF A DNA HELICASE WITH A 3' TO 5' POLARITY. REGC
CC UNWIND BRANCHED DUPLEX DNA (Y-DNA) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE REGC SUBFAMILY OF HELICASES.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute as long as its contents are used
CC without modification and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001159; AAC66942.1; --
CC TIGR; BB0581; J001410; --
CC InterPro: IPR001650; --
CC Pfam; PF00270; DEAD; 1.
CC Pfam; PF00271; helicase_C; 1.
CC Helicase; DNA repair; ATP-binding; DNA recombination; DNA-binding.
CC NE_BIND 290 297 ATP (POTENTIAL).
CC SITE 391 394 DEQH BOX
CC SEQUENCE 686 AA: 79043 MW; 5D4B20F38E6BA50B CRC64;
CC -----
Query Match 29.8%; Score 43.5; DB 1; Length 686;
Best Local Similarity 34.2%; Pred. No. 1.2e+02;
Matches 13; Conservative 3; Mismatches 7; Indels 15; Gaps 2;
OY 1 VEGG-----LYGVGPRAEPQTE-----ARRHYD 23
DB 292 VSGKTVIALLSGLPLIEAGYQVAFMAPTDLARQHYD 329
| | | | |
RESULT 27
SYLM_NEUCR SYLM_NEUCR STANDARD; PRT; 994 AA.
ID SYLM_NEUCR
AC P5181; 1990 (Rel. 14; Created)
DT 01-JUN-1990 (Rel. 14; Last sequence update)
DT 01-APR-1994 (Rel. 29; Last annotation update)
DE LEUCYL-TRNA SYNTHETASE, MITOCHONDRIAL PRECURSOR (EC 6.1.1.4)
DE (LEUCINE--TRNA LIGASE) (LEUCOS).
LEU 5.
OS Neurospora crassa. Ascomycota; Pezizomycotina; Sordariomycetes;
GN Sukaryota Fungi
OC Saccharimycota; Basidiomycota; Neurospora.
OX NCBI_taxid=5141;
[1]
SEQUENCE FROM N.A.
MEDLINE=90097874; Pubmed=2574823;
Chow C.M., Metzberg R.L., Rajbandary U.L.;
"Nuclear gene for mitochondrial leucyl-tRNA synthetase of Neurospora
crassa": Isolation, sequence characteristics, mapping, and evidence that
the gene encodes a precursor of the mature protein.
Mol Cell Biol 9:4631-4644(1989)
FL MO
-1- CATALYTIC ACTIVITY: ATP + L-LEUCINE + TRNA(LEU) = AMP +
PYROPHOSPHATE + L-LEUCYL-TRNA(LEU).
SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
-1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J30472; XAAB3599.1; --
CC TIGR; BB0581; J001410; --
CC InterPro: IPR001650; --
CC Pfam; PF00270; DEAD; 1.
CC Pfam; PF00271; helicase_C; 1.
CC Helicase; DNA repair; ATP-binding; DNA recombination; DNA-binding.
CC NE_BIND 290 297 ATP (POTENTIAL).
CC SITE 391 394 DEQH BOX
CC SEQUENCE 686 AA: 79043 MW; 5D4B20F38E6BA50B CRC64;

```


[illegible]

P18833; Q19359;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CUTICLE COLLAGEN 8.
GN COL-8 OR F11H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA MEDLINE=89326131; PubMed=2753356;
RA Cox G.N., Fields C., Kramer J.M., Rosenzweig B., Hirsh D.;
RT "Sequence comparisons of developmentally regulated collagen genes of
Caenorhabditis elegans.";
RT Gene 76:331-344(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Menezes S.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS.
CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
CC COLLAGENS. COL-7, COL-8 AND COL-19 BELONGS TO THE SAME GROUP OF
CC COLLAGEN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M25479; AAA27993.1; .
DR EMBL; U40187; AAA81159.1; .
DR PIR; JS0168; JS0168.
DR WormPep; F11H8.3; CE04376.
DR InterPro; IPR000087; .
DR Pfam; PF01391; Collagen; 2.
DR Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
KW DOMAIN 95 124 TRIPLE-HELICAL REGION.
FT DOMAIN 141 269 TRIPLE-HELICAL REGION.
FT CONFLICT 43 43 T -> R (IN REF. 2).
FT CONFLICT 142 142 C -> R (IN REF. 2).
FT CONFLICT 190 190 C -> R (IN REF. 2).
SQ SEQUENCE 282 AA; 27989 MW; 49ABBF66F9D3D32 CRC64;

Query Match 29.5%; Score 43; DB 1; Length 282;
Best Local Similarity 80.0%; Pred No. 57;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 PRAEPTGEAR 19
II IIIII I
Db 196 PRGEPCTEYR 205

RESULT 31
DAF_HUMAN
ID DAF_HUMAN STANDARD; PRT; 381 AA.
AC P08174; P09679;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COMPLEMENT DECAT-ACCELERATING FACTOR PRECURSOR (CD55).
GN DAF OR CD55.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87115845; PubMed=2433596;
RA Caras I.W., Davitz M.A., Rhee L., Weddell G., Martin D.W. Jr.,
RA Nussenzweig V.;
RT "Cloning of decay-accelerating factor suggests novel use of splicing
to generate two proteins.";
RL Nature 325:545-549(1987).
RN [2]
RP SEQUENCE OF 6-381 FROM N.A. (DAF-2 FORM).
RX MEDLINE=87175602; PubMed=2436222;
RA Medof M.E., Lublin D.M., Holers V.M., Ayers D.J., Getty R.R.,
RA Leykam J.F., Atkinson J.P., Tykocinski M.L.;
RT "Cloning and characterization of cDNAs encoding the complete sequence
of decay-accelerating factor of human complement.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2007-2011(1987).
RN [3]
RP SEQUENCE OF 1-100 FROM N.A.
RX MEDLINE=91271256; PubMed=1711208;
RA Ewulonu U.K., Ravi L., Medof M.E.;
RT "Characterization of the decay-accelerating factor gene promoter
region.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:4675-4679(1991).
RN [4]
RP SEQUENCE OF 35-46.
RC TISSUE-Urine;
RX MEDLINE=91291869; PubMed=1712233;
RA Nakano Y., Sugita Y., Ishikawa Y., Choi N.-H., Tobe T., Tomita M.;
RT "Isolation of two forms of decay-accelerating factor (DAF) from human
urine.";
RL Biochim. Biophys. Acta 1074:326-330(1991).
RN [5]
RP GPI-ANCHOR.
RX MEDLINE=91033238; PubMed=1824699;
RA Moran P., Raab H., Kohr W.J., Caras I.W.;
RT "Glycophospholipid membrane anchor attachment. Molecular analysis of
the cleavage/attachment site.";
RL J. Biol. Chem. 266:1250-1257(1991).
RN [6]
RP DISULFIDE BONDS IN SUSHI DOMAINS.
RX MEDLINE=92305034; PubMed=1377029;
RA Nakano Y., Sumida K., Kikuta N., Miura N.-H., Tobe T., Tomita M.;
RT "Complete determination of disulfide bonds localized within the short
consensus repeat units of decay accelerating factor (CD55 antigen).";
RL Biochim. Biophys. Acta 1116:235-240(1992).
RN [7]
RP FUNCTION AS A ECHOVIRUS RECEPTOR.
RX MEDLINE=95045399; PubMed=7525274;
RA Ward T., Pipkin P.A., Clarkson N.A., Stone D.M., Minor P.D.,
RA Almond J.W.;
RT "Decay-accelerating factor CD55 is identified as the receptor for
echovirus 7 using CELICS, a rapid immuno-focal cloning method.";
RL EMBO J. 13:5070-5074(1994).
RN [8]
RP VARIANT BLOOD GROUP DR(A-).
RX MEDLINE=94325573; PubMed=7519480;
RA Lublin D.M., Mallinson G., Poole J., Reid M.E., Thompson E.S.,
RA Ferdman B.R., Telen M.J., Anstee D.J., Tanner M.J.A.;
RT "Molecular basis of reduced or absent expression of
decay-accelerating factor in Cromer blood group phenotypes.";
RL Blood 84:1276-1282(1994).
CC -!- FUNCTION: THIS PROTEIN RECOGNIZES C4B AND C3B FRAGMENTS THAT
CC CONDENSE WITH CELL-SURFACE HYDROXYL OR AMINO GROUPS WHEN NASCENT
CC C4B AND C3B ARE LOCALLY GENERATED DURING C4 AND C3 ACTIVATION.
CC INTERACTION OF DAF WITH CELL-ASSOCIATED C4B AND C3B POLYPEPTIDES
CC INTERFERES WITH THEIR ABILITY TO CATALYZE THE CONVERSION OF C2 AND

RESULT	33
ID	MOXR_METEX
ID	MOXR_METEX STANDARD; * PRT; 341 AA.
CD	P30621;
DT	01-APR-1993 (Rel. 25, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-JUL-1998 (Rel. 36, Last annotation update)
DE	MOXR PROTEIN (MXAR PROTEIN).
DE	MOXR OR MXAR.
GN	Methylobacterium extorquens.
OS	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC	Methylobacterium group; Methylobacterium.
OX	NCBI_TaxID=408;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-AM1 / NCIB 9133;
RC	MEDLINE=97151731; PubMed=8997703;
RA	Amaratunga K., Goodwin P.M., O'Connor C.D., Anthony C.;
RT	"The methanol oxidation genes mxaFUGIR (S) ACKLD in Methylobacterium
RT	extorquens";
RL	FEMS Microbiol. Lett. 146:31-38(1997).
RN	[2]
RP	ERRATUM.
RC	MEDLINE=97306655; PubMed=9163922;
RA	Amaratunga K., Goodwin P.M., O'Connor C.D., Anthony C.;
RA	FEMS Microbiol. Lett. 150:175-177(1997).
RN	[3]
RP	SEQUENCE OF 1-81 FROM N.A.
RC	STRAIN-AM1 / NCIB 9133;
RC	MEDLINE=89350892; PubMed=2504152;
RA	Nunn D.N., Day D., Anthony C.;
RT	"The second subunit of methanol dehydrogenase of Methylobacterium
RT	extorquens AM1.";
RL	Biochem. J. 260:857-862(1989).
CC	-!- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF FORMATION OF ACTIVATED METHANOL DEHYDROGENASE.
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC	-!- SIMILARITY: WITH THE CORRESPONDING PROTEIN IN P. DENITRIFICANS.

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL021817; CAB40198.1; -.
DR DR EMBL; Z69240; CAA93240.1; -.
DR DR Hypothetical protein.
KW KW SEQUENCE 357 AA; 40709 MW; 5529B8D3B88D91A9 CRC64;
SQ
Query Match          29.1%; Score 42.5; DB 1; Length 357;
Best Local Similarity 37.1%; Pred. No. 84;
Matches 13; Conservative 3; Mismatches 8; Indels 11; Gaps 2;

OY      4 GLYGQGVPRAE--PGTEARRH-----YDEGVR 27
       |: | | | | | | | | | | | | | |
Db    82 GVNGIDRANETPGCIARRNLLRLRYETVDAGVR 116

```

```

RN      35
PKG_THETH
AC      01-MAR-1989 (Rel. 10, Created)
DT      01-MAR-1992 (Rel. 21, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
GN      PKG_SPHOGLYCERATE KINASE (EC 2.7.2.3)
OS      Thermus aquaticus (subsp. thermophilus).
OC      Bacteria; Thermus/Deinococcus group; Thermus.
OX      NCBI_TaxID=274;
RN      36
SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
MEDLINE=89025722; PubMed=3052437;
Bowen D., Littlechild J.A., Fothergill J.E., Watson H.C., Hall L.;
"Nucleotide sequence of the phosphoglycerate kinase gene from the
extreme thermophile Thermus thermophilus. Comparison of the deduced
amino acid sequence with that of the mesophilic yeast
phosphoglycerate kinase, (1988).
PLOCOCYB 3:51-55;
CC      1- CATALYTIC ACTIVITY: (ATP + 3-PROPHO-D-GLYCERATE = ADP +
CC      3-PROPHO-D-GLYCEROL PHOSPHATE.
CC      -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC      -1- SUBUNIT: MONOMER.
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC      -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: X12464; CAA31006.1;
CC      PIR: S01307; TVTWG.
CC      InterPro: IPR001576;
CC      Pfam: PF00162; PKG_1; KINASE.
CC      PRINIS: P00162; PKG_1; KINASE.
CC      PROSITE: PS00111; POLYMERASE KINASE; 1.
CC      Transferase; Kinase; Glycolysis.
CC      INT_MET 0
CC      SEQUENCE 389 AA; 41646 MW; 800F0535AB7F3C27 CRC64;
Query Match 29.1%; Score 42.5; DB 1; Length 389;
Best Local Similarity 37.5%; Pred No. 92;
Matches 12; Conservative 3; Mismatches 12; Indels 5; Gaps 1;
QY      1 VEGGLYQGVPRAE-----PGTEARRHYDEGVR 27
DB      72 VGEALRAHLPEARFAFPFGPGSEARREAEALR 103
RN      37
ECCL_HALEL
AC      Q2EUG; 2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      L-ECTOINE SYNTHASE.
GN      ECTC.
OS      Halomonas elongata.
OC      Bacteria; Proteobacteria; gamma subdivision; Halomonadaceae;
OX      NCBI_TaxID=2746;
RN      38
SEQUENCE FROM N.A.
RN      39
SEQUENCE FROM N.A.
STRAIN-ATCC 13059 / AS019;
MEDLINE=88232419; PubMed=2836698;
Yeh P., Sicard A.M., Sinskey A.J.;
"Nucleotide sequence of the lysA gene of Corynebacterium glutamicum
and possible mechanism of its regulation of its expression."
Mol. Gen. Genet. 212:112-119 (1988).
RN      40
SEQUENCE OF 1-49 FROM N.A.
STRAIN-ATCC 13059 / AS019;
MEDLINE=91186817; PubMed=2082143;
Marcel T., Archer J.A.C., Mengin-Lecomte J.;
"Nucleotide sequence of the glutamicum lysA gene."
Mol. Microbiol. 4:1819-1830 (1990).
RN      41
CATALYTIC ACTIVITY: MESO-2,6-DIAMINOHEPTANEDIOATE = L-LYSINE +
CO(2).
CC      -1- COFACTOR: PYRIDOXAL PHOSPHATE
CC      -1- PATHWAY: LAST STEP IN THE SYNTHESIS OF LYSINE.
CC      -1- REACTION: LAST STEP IN THE SYNTHESIS OF LYSINE.
CC      -1- REACTION ARE CRITICAL FOR CELL GROWTH.
CC      -1- SIMILARITY: BELONGS TO FAMILY 2 OF ORNITHINE, DAP, AND ARGININE
CC      DECARBOXYLASES.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: X04760; CAA30442.1;
CC      PIR: S03827; S03827.
CC      InterPro: IPR000183;
CC      InterPro: IPR002986;
CC      Pfam: PF002778; Orn_DAP_Arg_dec; 1.
CC      PRINIS: P03827; ORN_DAP_ARG_DEC.
CC      PROSITE: PS00878; ODR_DC_2.1; 1.
CC      PROSITE: PS00878; ODR_DC_2.2; 2.
CC      Lysine biosynthesis; Lyase; Decarboxylase; Pyridoxal phosphate.
CC      BINDING 75 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC      SEQUENCE 445 AA; 47411 MW; F7F49A23EAG6CAB3 CRC64;
Query Match 29.1%; Score 42.5; DB 1; Length 445;
Best Local Similarity 52.4%; Pred No. 11e+02;
Matches 11; Conservative 1; Mismatches 8; Indels 1; Gaps 1;
QY      2 GEGLYQGV-PRAEPTGTEARRH 21
DB      154 GEGKIDVLIIVKFGIEANTH 174
RN      37
ECCL_HALEL
ID      ECCL_HALEL
STANDARD; PRT; 130 AA.
AC      Q2EUG; 2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      L-ECTOINE SYNTHASE.
GN      ECTC.
OS      Halomonas elongata.
OC      Bacteria; Proteobacteria; gamma subdivision; Halomonadaceae;
OX      NCBI_TaxID=2746;
RN      38
SEQUENCE FROM N.A.

```

```
RC STRAIN-DSM 3043;
RX MEDLINE-99123891; PubMed-9924816;
RA Canovas D., Vargas C., Calderon M.I., Ventosa A., Nieto J.J.;
RT "Characterization of the genes for the biosynthesis of the compatible
RT solute ectoine in the moderately halophilic bacterium Halomonas
RT elongata DSM 3043.";
RL Syst. Appl. Microbiol. 21:487-497(1998).
CC -!- FUNCTION: CYCLIC CONDENSATION OF GAMMA-N-ACETYL-ALPHA, GAMMA-
CC DIAMINO BUTYRIC ACID (ADABA) TO ECTOINE.
CC -!- PATHWAY: LAST STEP IN THE BIOSYNTHESIS OF ECTOINE (1,4,5,6-
CC TETRAHYDRO-2-METHYL-4-PYRIMIDINE CARBOXYLIC ACID).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ011103; CAA09485.1; -; EE0E0B30B1E761BE CRC64;
SQ SEQUENCE 130 AA; 14824 MW; EE0E0B30B1E761BE CRC64;

Query Match 28.8%; Score 42; DB 1; Length 130;
Best Local Similarity 42.1%; Pred. No. 36;
Matches 8; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 4 GLYQGVPRAEPTGEARRHY 22
Db 36 GFSFNTRIHPGTETHY 54
| : | | | | |
| : | | | | |

RESULT 38
YQ33_CAEEL STANDARD; PRT; 283 AA.
AC Q09233;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PUTATIVE CUTICLE COLLAGEN C09G5.3.
GN C09G5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC STRAIN-BRISTOL N2;
RA Palmer S.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
CC COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z46791; CAA86756.1; -;
DR WormPep: C09G5.3; CE01483.
DR InterPro: IPR000087; -;
DR InterPro: IPR002486; -;
DR Pfam: PF01484; Col_cuticle_N; 1.
```

```
DR Pfam: PF01391; Collagen; 2.
KW Hypothetical protein; Cuticle; Connective tissue; Repeat;
KW Multigene family; Collagen.
FT DOMAIN 94 122 TRIPLE-HELICAL REGION.
FT DOMAIN 139 201 TRIPLE-HELICAL REGION.
FT DOMAIN 204 269 TRIPLE-HELICAL REGION.
SQ SEQUENCE 283 AA; 28271 MW; 539DEF1392989D50 CRC64;

Query Match 28.8%; Score 42; DB 1; Length 283;
Best Local Similarity 41.9%; Pred. No. 79;
Matches 13; Conservative 4; Mismatches 8; Indels 6; Gaps 2;

QY 1 VGEGLQGVPRAEPTGEARRHYDEGV 26
Db 203 IGEAGPKGPCCA-FGTDGRCGNGTGDG 232
| : | | | | | | | | | | | | | |
| : | | | | | | | | | | | | | |

RESULT 39
CITG_KLEPN STANDARD; PRT; 302 AA.
AC P45414;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE CITG PROTEIN.
GN CITG.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN-ATCC 13882;
RX MEDLINE-95131756; PubMed-7830578;
RA Bott M., Dimroth P.;
RT "Klebsiella pneumoniae genes for citrate lyase and citrate lyase
RT ligase: localization, sequencing, and expression.";
RL Mol. Microbiol. 14:347-356(1994).
CC -!- FUNCTION: PROBABLY INVOLVED IN THE BIOSYNTHESIS OF THE PROSTHETIC
CC GROUP (2-(5''-PHOSPHORIBOSYL)-3'-DEPHOSPHOCOENZYME-A) OF THE GAMMA
CC SUBUNIT OF CITRATE LYASE.
CC -!- SIMILARITY: BELONGS TO THE CITG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X79817; CAA56218.1; -;
DR InterPro: IPR002736; -;
DR Pfam: PF01874; Citg; 1.
SQ SEQUENCE 302 AA; 32645 MW; 25AAD07DC3580056 CRC64;

Query Match 28.8%; Score 42; DB 1; Length 302;
Best Local Similarity 45.0%; Pred. No. 84;
Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 LYQGVPRAEPTGEARRHYDE 24
Db 250 LWQGGVLADGGLEALRQFDD 269
| : | | | | | | | | | | | | | |
| : | | | | | | | | | | | | | |

RESULT 40
FLAA_BORBU STANDARD; PRT; 341 AA.
ID FLAA_BORBU
AC P70856; O51612; Q44876;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
```



```
OS Arabidopsis thaliana (Mouse-ear cress).
OC Magnoliopsida: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:
OC Brassicales: Brassicaceae: Arabidopsis.
OX NCBI_TaxID=3702;
RN SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA K06176; PubMed=7579169;
RA Kinoshita T, Nishimura M, Hara Nishimura I.;
RT "Homologues of a vacuolar processing enzyme that are expressed in
RL different organs of Arabidopsis thaliana.";
RL Plant Mol. Biol. 29:81-89(1995).
CC CC
CC -1- FUNCTION: ASPARAGINE-SPECIFIC ENDOPEPTIDASE INVOLVED IN THE
CC PROCESSING OF VACUOLAR SEED PROTEIN PRECURSORS INTO THE MATURE
CC FORM.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ROSETTE LEAVES, CAULINE LEAVES
CC AND STEMS. NOT EXPRESSED IN THE SILIQUES.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C13; ALSO KNOWN AS THE
CC HEMOGLOBINASE FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the European Bioinformatics Institute and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D61393; BRA09614.1;
CC MEROPS: C13.002;
CC InterPro: IPR001096;
CC Pfam: PF01650; Peptidase_C13; 1.
CC PRINTS: PR00776; HEMOGLOBINASE.
CC SIGNAL: MULTIGENE FAMILY.
CC HYDROLASE: THIOLESTERASE; SIGNAL.
CC ACT_SITE 21 478
CC CHAIN 21 478
CC VACUOLAR PROCESSING ENZYME, ALPHA-
CC ISOZYME
CC SEQUENCE 478 AA; 52680 MW; B79497621FF30F74 CRC64;

Query Match 28.8%; Score 42; DB 1; Length 478;
Best Local Similarity 40.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 GRGQGVPRAPRGTE 17
DB 107 GEDVYNGVPKDTGDE 122

RESULT 45
ID DHE3_MOUSE
AC P26443;
DT 01-AUG-1992 (rel. 23, Created)
DT 01-AUG-1992 (rel. 23, Last sequence update)
DE GRGQGVPRAPRGTE 17
DE GLUTAMATE DEHYDROGENASE PRECURSOR (EC 1.4.1.3) (GDH).
GN GLUD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Brain;
RA MEDLINE=91274358; PubMed=1711373;
RA Tzimagiorgis G., Moschonas N.K.;
RT "Molecular cloning, structure and expression analysis of a
RT full-length mouse brain glutamate dehydrogenase cDNA.";
RL Plant Mol. Biol. 29:81-89(1995).
CC CC
CC -1- CATALYTIC ACTIVITY: L-GLUTAMATE + H(2)O + NAD(P)(+) =
CC -2- OXOGLOUTARATE + NH(3) + NAD(P)H.
```

```
CC CC
CC -1- SUBUNIT: HOMOHENANER.
CC -1- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the European Bioinformatics Institute and the EMBL outstation -
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X57024; CAA40341.1;
CC PIR: S16239; S16239.
CC SSF: PR03319; P2GM.3; MOUSE.
CC TIGR: M10007; P2GM.3; MOUSE.
CC MGD_MGI: 95753 P2Glud.
CC InterPro: IPR001625;
CC Pfam: PF00208; GLFV_dehydrog; 1.
CC PRINTS: PR00082; GLFV_DEHYDROG.
CC PROSITE: PS00074; NADP5_Mitochondrial_NADPH-dependent
CC KXICREDUCTASE; NADP5_Mitochondrial_NADPH-dependent
CC CHAIN 54 558
CC BINDING 84 84
CC ACT_SITE 183 183
CC SEQUENCE 558 AA; 61336 MW; 92738AA5A133838A CRC64;

Query Match 28.8%; Score 42; DB 1; Length 558;
Best Local Similarity 48.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 7; Mismatches 7; Indels 4; Gaps 2;

OY 4 GLYQGVPRAPRGTE--ARRHYDE 24
DB 31 GWARGQSAAPQGLTPVARRHYSE 55

Search completed: June 28, 2001, 11:57:29
Job time: 270 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:55:36 ; Search time 72.61 Seconds
(without alignments)
49.198 Million cell updates/sec

Title: US-09-439-313-566

Perfect score: 146

Sequence: 1 VGEGLYQGVPRAPETGTEARRHYDEGVR 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database :

1: SPTREMBL16.*

2: sp_archaea.*

3: sp_bacteria.*

4: sp_fungi.*

5: sp_human.*

6: sp_invertebrate.*

7: sp_mammal.*

8: sp_mhc.*

9: sp_organelle.*

10: sp_phage.*

11: sp_plant.*

12: sp_rodent.*

13: sp_unclassified.*

14: sp_vertebrate.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	ID	Description
1	69	47.3	599	5	Q9VSV1	Q9VSV1 drosophila
2	63	43.2	742	4	Q94841	Q94841 homo sapien
3	63	43.2	742	6	Q29397	Q29397 bos taurus
4	63	43.2	742	11	Q9JIS5	Q9JIS5 mus musculus
5	55	37.7	533	4	Q9Y2W3	Q9Y2W3 homo sapien
6	51	34.9	184	1	Q9HKH4	Q9HKH4 thermoplasma
7	51	34.9	241	5	Q9WS12	Q9WS12 drosophila
8	50.5	34.6	610	2	Q91231	Q91231 pseudomonas
9	49.5	33.9	495	2	O52209	O52209 serratia ma
10	49.5	33.9	665	2	Q48373	Q48373 anthinobac
11	49	33.6	121	5	Q9NHM5	Q9NHM5 plasmodium
12	49	33.6	420	2	Q9RKQ2	Q9RKQ2 streptomyce
13	49	33.6	2314	2	O69822	O69822 streptomyce
14	48	32.9	153	2	O86586	O86586 streptomyce
15	48	32.9	356	2	Q9K3R1	Q9K3R1 streptomyce
16	48	32.9	482	10	Q9XFZ5	Q9XFZ5 vigna mungo
17	48	32.9	616	2	Q49182	Q49182 mycobacteri
18	48	32.9	729	13	Q9PVF7	Q9PVF7 brachydanio
19	48	32.9	808	2	O70021	O70021 saccharopol

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	599 AA.
Q9VSV1	Q9VSV1			
AC	Q9VSV1			
DT	01-MAY-2000 (TREMREL. 13, Created)			
DT	01-MAY-2000 (TREMREL. 13, Last sequence update)			
DT	01-JUN-2000 (TREMREL. 14, Last annotation update)			
DE	CG4484 PROTEIN.			
GN	CG4484.			

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

Q40848 picea glauc
Q56154 streptomyce
Q9pbj5 xylella fas
Q56434 thermus aqu
Q9k3m0 streptomyce
Q9mf80 beta vulgar
Q9v077 pyrococcus
Q9tqg2 drosophila
Q96660 drosophila
Q82265 arabidopsis
Q91p20 arabidopsis
Q9xgm9 arabidopsis
Q9ru55 deinococcus
Q9n3r9 caenorhabdi
Q9ue24 homo sapien
Q65360 orgyia pseu
Q9udq4 homo sapien
Q9r7s4 escherichia
Q9uza5 pyrococcus
Q9ulu3 homo sapien
Q9f584 agrobacteri
Q9z206 caenorhabdi
Q45395 bordetella
Q9w077 drosophila
Q916t0 salmonella
Q66027 mycobacteri
Q22273 arabidopsis
Q65227 arabidopsis
Q9shu0 arabidopsis
Q9v5a6 drosophila
Q9r230 rattus norv
Q46486 corynebacte
P90273 human immun
Q9gl19 cynocephalu
Q9kyt6 streptomyce
Q9rl88 pseudomonas
O88746 mus musculu
P76523 escherichia
Q9hr25 halobacteri
Q9hmn4 halobacteri
Q9kyr2 streptomyce
Q35906 mus musculu
Q9yfi4 aeropyrum p
Q9lqz2 oryza sativ
P95236 mycobacteri
Q9y9e8 aeropyrum p

RA Janz R., Goda Y., Geppert M., Missler M., Sudhof T.C.;
RT "SV2A and SV2B function as redundant Ca2+ regulators in
neurotransmitter release."
RL Neuron 24:1003-1016(1999).
DR EMBL; AF196781; RAF87321.1;
DR EMBL; AF196780; RAF87321.1; JOINED.
DR InterPro; IPR001066;
DR Pfam; PF00083; sugar_tr.1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN.1.
SQ SEQUENCE 742 AA; 82647 MW; 1074857FD13ED894 CRC64;

Query Match 43.2%; Score 63; DB 11; Length 742;
Best Local Similarity 51.7%; Pred. No. 1.1;
Matches 15; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 3 EGLYQGVPRAEPTGTEARRHYD----EGVR 27
II :||:||||| : : : |||
DB 95 EGEYQGIPTAESGGKGERMADGAPLAGVR 123

RESULT 5
QY2W3 PRELIMINARY; PRT; 533 AA.
ID QY2W3;
AC QY2W3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE DNB-5 (FRAGMENT).

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RA Amler L.C., Bauer A., Corvi R., Dihlmann S., Praml C., Cavenee W.K.,
RA Schwab M., Hampton G.M.;
RT "Identification and Characterization of Novel Genes Located at the
RT t(1;15)(p36.2;q24) Translocation Breakpoint in the Neuroblastoma Cell
RT Line NCP."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF118274; AAD27583.1;
FT NON_TER
SQ SEQUENCE 533 AA; 57048 MW; 6AF8BA162FC9D1C8 CRC64;

Query Match 37.7%; Score 55; DB 4; Length 533;
Best Local Similarity 42.3%; Pred. No. 11;
Matches 11; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 1 VCEGLYQGVPRAEPTGTEARRHYDEGV 26
:||:||||| : : : |||
DB 333 MGEVVFQDPKAPHTSEAYQKYNQSV 358

RESULT 6
Q9HKH4 PRELIMINARY; PRT; 184 AA.
ID Q9HKH4;
AC Q9HKH4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN TA0625.

GN TA0625.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC Thermoplasma
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
The genome sequence of Drosophila melanogaster.;

RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL; AL445064; CAC11764.1;
KW Hypothetical protein.
SQ SEQUENCE 184 AA; 20720 MW; 6C416E059543D0F CRC64;

Query Match 34.9%; Score 51; DB 1; Length 184;
Best Local Similarity 36.0%; Pred. No. 13;
Matches 9; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 3 EGLYQGVPRAEPTGTEARRHYDEGVR 27
||:||||| : : : |||:|
DB 11 EGRWQINDSDIPEESLRHEERIR 35

RESULT 7
Q9W5I2 PRELIMINARY; PRT; 241 AA.
ID Q9W5I2;
AC Q9W5I2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG17683 PROTEIN.

GN CG17683.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu P., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M., Pan S.,
RA Palazzotto M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
The genome sequence of Drosophila melanogaster.;

```

SEQUENCE FROM N.A. MEDLINE=7509689; PubMed=1089756;
Hedges R.W., Rodriguez-Lemoine V., Datta N.;
"R factors from Serratia marcescens.";
J. Gen. Microbiol. 86:88-92(1975).
[2] SEQUENCE FROM N.A. PubMed=8366028;
MEDLINE=9328935; Levine A.S., Woodgate R.;
H.C. Killaeva O.I. Levine A.S., Woodgate R.;
"A rapid method for cloning mutagenic DNA repair genes: isolation of
umu-complementing clones from multidrug resistance plasmids R391,
R446b, and R471a.";
J. Bacteriol. 175:5411-5419(1993).
[3] SEQUENCE FROM N.A. Ewing E.V., Wootton J.C., Levine A.S., Woodgate R.;
MultiRes. 0:0-0(1997).
MUT - SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE
TRANSCRIPTASE).
EMBL: AF027768; AAC82519.1; -
InterPro: IPR000477; -
Pfam: PF00078; tvt: l...ase.
RNA predicted DNA pol...ase.
SEQUENCE 495 AA; 57746 MW; ECA93A980D961A15 CRC64;

Query Match 33.9%; Score 49.5; DB 2; Length 495;
Best Local Similarity 61.1%; Pred. No.59;
Matches 11; Conservative 2; Mismatches 2; Indels 3; Gaps

QY 5 LYGVPRAPGCTEARRHY 22
I...II...II...II...
DB 120 LHQGSYRAOPG--RRHY 134

RESULT 10
ID C48373 PRELIMINARY; PRT; 665 AA.
QC C48373
DT 01-NOV-1996 (Tremblrel. 01, Created)
DI 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE CHITININASE PRECURSOR.
CHI 69.
OS Bacterium lividum.
ON panthobacterium lividum.
OC Tanninobacterium.
OC Tanninobacterium.
NCBI_TaxID=29581.
[1] SEQUENCE FROM N.A.
MEDLINE=96013069; PubMed=7557339;
Gleave A.P., Taylor R.K., Morris B.A., Greenwood D.R.;
Chitoning and sequencing the 69-kDa extracellular
chitinase from Bacterium lividum.";
FEMS Microbiol. Lett. 131:279-288(1995).
RL - SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL: U07025; AAA83223.1; -.
DR HSP; P07254; ICTN.
DR InterPro: IPR000529; -
DR InterPro: IPR001579; -
DR InterPro: IPR003610; -
DR Pfam: PF00704; Glyco_hydro_18; 2.
DR PROSITE; PS01095; CHITINASE_18; 1.
DR PROSITE; PS00761; SPASE_I_3; UNKNOWN_1.
DR SMART; SM00495; ChitBd3; 1.
KW Glycosidase; Hydrolase; Signal.
KW POTENTIAL.
FT CHAIN 24 665
SO SEQUENCE 665 AA; 69748 MW; FCF371332CIC6839 CRC64;

```


ID Q96596 PRELIMINARY; PRT; 153 AA.
 AC Q96596;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE PUTATIVE LIPOPROTEIN.
 GN SC2H4.03.
 GN Streptomyces coelicolor.
 OS Streptomyces coelicolor.
 OC Actinomycetales; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Oliver K., Harris D.; to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RN [3]
 RN Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
 RN [3]
 RN MEDLINE=97000351; PubMed=8843436;
 RA Kinashi H., Hopwood D.A.;
 RN [3]
 RN "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 DL EMBL; AL031514; CAA20597.1;
 SQ SEQUENCE 153 AA; 15655 MW; 695EA9DD6847BDA8 CRC64;

Query Match 32.9%; Score 48; DB 2; Length 153;
 Best Local Similarity 32.9%; Pred. NO. 27;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 2 GEGLYGVPRAPPGTE 17

DB 92 GDGLYMTSPRARGSE 107

RESULT 15

ID Q9K3R1 PRELIMINARY; PRT; 356 AA.
 AC Q9K3R1;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE PUTATIVE ALDOLASE.
 GN 2SCG4.03C
 GN Streptomyces coelicolor.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RN [2]
 RN Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RN [3]
 RN Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
 RN [3]
 RN Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for
 RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL; AL360034; CAB95974.1;
 SQ SEQUENCE 356 AA; 39108 MW; 9BC43F91B4ACAF22 CRC64;

Query Match 32.9%; Score 48; DB 2; Length 356;
 Best Local Similarity 64.3%; Pred. NO. 68;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 14 PGTEARRHYDEGVR 27

DB 4 PKIDARRHHDEVR 17

RESULT 16

ID Q9XFZ5 PRELIMINARY; PRT; 482 AA.
 AC Q9XFZ5;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE ASPARAGINYL ENDOPEPTIDASE (VMEPE-1A).
 OS Vigna mungo (Rice bean) (Black gram).
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OS Magnoliophyta; Eudicotyledons; Rosidae; eucotyledons; Rosidae; eucotyledons; Papilionoideae; Vigna.
 OX NCBI_TaxID=3915;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=COTYLEDON;
 RN MEDLINE=99178794; PubMed=10080709;
 RA "Kamota T., Minamikawa T., and Kato T. (1999) Cloning and characterization of Vigna mungo processing
 RT enzyme 1 (VMEPE-1) an asparaginyl endopeptidase possibly involved in
 RT post-translational processing of a vacuolar cysteine endopeptidase
 RT (SH-EP).";
 RL Plant Mol. Biol. 39:63-73(1999).
 DR EMBL; D89972; BAA76745.1;
 DR InterPro; IPR001519;
 DR Pfam; PF01650; Peptidase_G13; 1.
 DR PRINTS; PR00776; HEMOGLOBINASE.
 DR PRODOM; PD000971; 1.
 SQ SEQUENCE 482 AA; 53211 MW; C8262E868ED378E8 CRC64;

Query Match 32.9%; Score 48; DB 10; Length 482;
 Best Local Similarity 40.0%; Pred. NO. 93;
 Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 2 GEGLYGVPRAPPGTEARRH 21

DB 111 GDVYGVPRDYTGEDATAH 130

RESULT 17

ID Q49182 PRELIMINARY; PRT; 616 AA.
 AC Q49182;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE DNA SEQUENCE, ORFS 1 AND 2.
 OS Mycobacterium fortuitum.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1766;
 RX [1]
 RP SEQUENCE FROM N.A.
 RN MEDLINE=92311251; PubMed=1615063;
 RA Labidi A., Mardis E., Roe B.A., Wallace R.J. Jr.;

```

RA  Gaiser S., Boehm G.A., Doumith M., Raynal M.C., Dhillon N.,
RA  Cortes J., Leadlay P.F.;
RL  Mol. Gen. Genet. 0:0-0(0).
DR  EMBL: Y14327; CAA74702.1; -.
DR  InterPro: IPR001764; -.
DR  InterPro: IPR002772; -.
DR  Pfam: PF00933; Glyco_hydro_3; 1.
DR  Pfam: PF01915; Glyco_hydro_3_C; 1.
DR  PRINTS: PR00133; GLHYDLASE3.
SQ  SEQUENCE 808 AA; 86677 MW; 0C619016CC00A751 CRC64;

Query Match 32.9%; Score 48; DB 2; Length 808;
Best Local Similarity 42.1%; Pred. No. 1.6e+02;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 8 GVPRAEPCTEARRHYDEGV 26
Db 651 GPARFFGVGKGVHISEGI 669
      | | | | | : : | | | :
      | | | | | : : | | | :

RESULT 20
Q40848 PRELIMINARY; PRT; 175 AA.
ID Q40848 AC Q40848;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE LATE EMBRYOGENESIS ABUNDANT PROTEIN.
GN EMB23.
OS Picea glauca (White spruce).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Coniferopsida; Coniferales; Pinaceae; Picea.
OX NCBI_TaxID=3330;
RN [1]
RP SEQUENCE FROM N.A.
RA Dong J.Z., Dunstan D.I.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBDJ databases.
DR EMBL: L47603; AAB01558.1; -.
DR Mendel; 12550; Picol; 1750; 12550.
SQ SEQUENCE 175 AA; 19128 MW; 0CD8BD4B3F80BA0B CRC64;

Query Match 32.5%; Score 47.5; DB 10; Length 175;
Best Local Similarity 48.0%; Pred. No. 37;
Matches 12; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

QY 1 VGEGLYQGVPR-AEPGTEARRHYDE 24
Db 114 LSEGAATVERDADPHVEVRREYDE 138
      : | | | | | : | | | | |
      : | | | | | : | | | | |

RESULT 21
Q56154 PRELIMINARY; PRT; 234 AA.
ID Q56154 AC Q56154;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE JADR1.
GN JADR1.
OS Streptomyces violaceus (Streptomyces venezuelae).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1936;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISP5230;
RX MEDLINE=96042086; PubMed=7592375;
RT Yang K., Han L., Vinling L.C.;
RT "Regulation of fadomycin B production in Streptomyces venezuelae
RT ISP5230: involvement of a repressor gene, jadR2.";
RL J. Bacteriol. 177:6111-6117(1995).

```


-1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
CC -1- SIMILARITY: TO BACTERIAL REGULATORY PROTEINS INVOLVED IN SIGNAL
CC TRANSDUCTION.

EMBL: U24659; AAB36584.1; --
DR HSSP: P08402; 1B00.
DR InterPro: IPR001789; --
DR InterPro: IPR001867; ase_reg; 1.
DR Pfam: PF00486; trans_reg_C; 1.
DR SMART: SM00448; REC; 1.
DR DNA-binding; Phosphorylation; Sensory transduction;
KW Transcription regulation.
SQ SEQUENCE 234 AA; 25606 MW; CC251108FEC3906D CRC64;

Query Match 32.5%; Score 47.5; DB 2; Length 234;
Best Local Similarity 43.3%; Pred. No. 51;
Matches 13; Conservative 2; Mismatches 8; Indels 7; Gaps 1;

OY 2 GEGLYQGVVPRAPGTEARRHYDE 24
DB 30 GEGLYQGVVPRAPGTEARRHYDE 24

RESULT 22

ID O9PBJ5 PRELIMINARY; PRT; 127 AA.
AC O9PBJ5, 2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE APAG PROTEIN.
GN XE2149.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.G., Alvarado C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Frega J.S., Franca S.C., Franco M.M., Gomes S.L., Gruber A.,
RA Ho P.T., Hohnselt J.D., Junqueira M.L.S., Kemp E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro V.C.B.,
RA Moon H.H., Negreiros F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.R., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M., Siqueira D., Tsiang S., Tsubota A.,
RA Vallada H., Van Sluys M.A., Varjoski-Almeida S.M., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AE004029; AAF84948.1; -- 3DD5206C2937BE25 CRC64;
SQ SEQUENCE 127 AA; 14247 MW; 3DD5206C2937BE25 CRC64;

Query Match 32.2%; Score 47; DB 2; Length 127;
Best Local Similarity 48.0%; Pred. No. 31;
Matches 12; Conservative 1; Mismatches 10; Indels 1; Gaps 1;

OY 2 GEGLYQGVVPRAPGTEARRHYDE 26
DB 66 GEGLYQGVVPRAPGTEARRHYDE 26

RESULT 23

ID O56434 PRELIMINARY; PRT; 134 AA.
AC O56434;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE RIBOSOMAL PROTEIN L11 (FRAGMENT).
GN RPL11.
OS Thermus aquaticus (subsp. thermophilus).
OC Bacteria; Firmicutes; Streptococcaceae; Streptomyces group; Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94124036; PubMed=8294036;
RA Heinrich T., Erdmann V.A., Hartmann R.K.;
RT "Sequence of the gene encoding ribosomal protein L11 from Thermus
RA thermophilus." Hb.
RL Gene 136:373-374(1993).
DR EMBL: L10371; AAA27503.1; --
DR Ribosomal protein.
FT NON_TER 1
SQ SEQUENCE 134 AA; 14336 MW; C8EB750B7B75EFC CRC64;

Query Match 32.2%; Score 47; DB 2; Length 134;
Best Local Similarity 41.7%; Pred. No. 33;
Matches 10; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

OY 2 GEGLYQGVVPRAPGTEARRHYDE 25
DB 40 GEGLYQGVVPRAPGTEARRHYDE 25

RESULT 24

ID O9K3M0 PRELIMINARY; PRT; 216 AA.
AC O9K3M0, 2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE DNA REPAIR PROTEIN.
GN SCG20A.20C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RL Nature 406:151-159(2000).
DR EMBL: AE004029; AAF84948.1; -- 3DD5206C2937BE25 CRC64;
SQ SEQUENCE 127 AA; 14247 MW; 3DD5206C2937BE25 CRC64;

OS *Pyrococcus abyssi*.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248285; CAB49828.1; -

DE SMOX OR SMAD2 OR CG2262.
 GN Drosophila melanogaster (Fruit fly).
 OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Muscomor
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomor
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA ADAMS M. D., Ceiniker S. E., Holt R. A., Evans C. A., Gocayne J. D.,

RA Ananathides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards M.D., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers Y.H., C. Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayan A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck H., Brockstein D., Center A.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fioder A., Gong F., Goran J., Garg N.S., Galbart W.M., Glasser K.,
RA Glisic N., L. L. D. Helina, T. J. Hernandez, J. R. Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson J.,
RA McKusick G., Mohlman B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
RA Shue B.C., Sinden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,
RA Swirsky R., Tector C., Turley K.C., Wei M.-H., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster."
RA [Genome 287:2185-2195 (2000)].
RP SEQUENCE FROM N.A.
RA Brummel T., Abdollah S., Haery T., Shimell M.J., Merriam J.,
RA Raftery L., Wrana J.L., O'Connor M.B.;
RT "The Drosophila activin receptor baboon signals through Dmad2 and
RT controls cell proliferation but not patterning during larval
RT development."
RA [3] Genes Dev. 0:0-0 (1998).
RP SEQUENCE FROM N.A.
RA Das P., Inoue H., Baker J.C., Beppu H., Kawabata M., Harland R.M.,
RA Miyazono K., Padgett R.W.;
RT Drosophila dmad2 and Atr-1 comprise part of the first functional
RT invertebrate activin receptor pathway.
RA [Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.]
DR EMBL: AF003443; AAF46330.1; -
DR EMBL: AF101386; AAD11458.1; -
DR EMBL: AF109132; AAD22443.1; -
DR Flybase: FBgn0025800; Smox.
DR InterPro: IPR001132; -
DR InterPro: IPR001132; -
DR SMART: SM00523; DWA; 1.
DR SMART: SM00523; DWA; 1.
RP SEQUENCE 486 AA; 53676 MW; 2BF312FD13534F98 CRC64;

Query Match 32.2%; Score 47; DB 5; Length 486;
Best Local Similarity 45.8%; Pred. No. 13e+02; Indels 0; Gaps 0;
Matches 11; Conservative 2; Mismatches 11;
OY 4 GLYQGVPRAEPTGTEARRHYDEGVR 27
DB 333 GLLSNVNRNEVVEQTRRHIGKGRV 356

RESULT 29

082265
ID 082265 PRELIMINARY; PRT: 921 AA.
AC 082265;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DE 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE PUTATIVE NUCLEAR PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1] JOURNAL FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA Shen M., Rensing C.M., Fraser C.M., Somerville C.R., Venter J.C.,
RA "Arabidopsis thaliana chromosome II BAC F17A22 genomic sequence."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RW EMBL: AC005309; AAC63652.1;
RP SEQUENCE 921 AA; 105737 MW; A6F5B95036F7CB3F CRC64;
Query Match 32.2%; Score 47; DB 10; Length 921;
Best Local Similarity 50.0%; Pred. No. 2.6e+02; Indels 0; Gaps 0;
Matches 9; Conservative 3; Mismatches 0;
OY 7 QGVPRAEPTGTEARRHYDE 24
DB 857 RGRPRKRPETKRLFE 874
RESULT 30
ID 09LP20 PRELIMINARY; PRT: 962 AA.
AC 09LP20;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE F14D7.6 PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1] JOURNAL FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Liu S.X., Chan C., Sakano H., Yu G., Lee J.M., Lenz C., Pham P.,
RA Toriumi M., Chin C., Chio J., Choi E., Chung M., Gonzalez A.,
RA Hong B., Liu A., Vaysberg M., Altafi H., Brooks S., Buehler E.,
RA Chao Q., Conn L., Conway A.B., Hansen N.F., Johnson-McPson Shinn P.,
RA Chan S., Kim C., Lam B., Madan A., Meeker J.R., Theologis A.,
RA Yu G., Chick A., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.,
RT "The sequence of BAC F14D7 from Arabidopsis thaliana chromosome 1."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;

RL Theologis;
DR Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC021198; AAF79886.1; -
DR TIGR; DRI539; -
KW Hypothetical protein.
SQ SEQUENCE 962 AA; 108167 MW; 0454713CBD77280A CRC64;

Query Match 32.2%; Score 47; DB 10; Length 962;
Best Local Similarity 70.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEGLYQGV 10 :|||:|
DB 820 IGDGLYGV 829 :|||:|

RESULT 31
Q9XGM9 ID Q9XGM9 PRELIMINARY; PRT; 1098 AA.
AC Q9XGM9;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE STROMALIN PROTEIN.
GN SA.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Valdeolmillos A.; Barbero J.;
RT "Molecular Cloning and expression of stromalin protein from
RT Arabidopsis thaliana."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RE EMBL; A17242965; CAB45374.1; -
SQ SEQUENCE 1098 AA; 125758 MW; 293ACEAD3D335471A CRC64;

Query Match 32.2%; Score 47; DB 10; Length 1098;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 7 QGVPRAPGTEARRHYDE 24 :|||:|
DB 1034 RGRPRKPETERKRLPDE 1051 :|||:|

RESULT 32
Q9RU55 ID Q9RU55 PRELIMINARY; PRT; 110 AA.
AC Q9RU55;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 11.3 KDA PROTEIN.
GN DR1539.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX White O.; Eisen J.A.; Heidelberg J.F.; Hickey E.K.; Peterson J.D.;
RA Dodson R.J.; Haft D.H.; Gwinn M.L.; Nelson W.C.; Richardson D.L.;
RA Moffat K.S.; Qin H.; Jiang L.; Pamphile W.; Crosby M.; Shen M.;
RA Vamathevan J.J.; Lam P.; McDonald L.; Utterback T.; Zalewski C.;
RA Makarova K.S.; Aravind L.; Daly M.J.; Minton K.W.; Fleischmann R.D.;
RA Ketchum K.A.; Nelson K.E.; Salzberg S.; Smith H.O.; Venter J.C.;
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."

RL Science 286:1571-1577(1999).
DR EMBL; A5001998; AAF11112.1; -
DR TIGR; DRI539; -
KW Hypothetical protein.
SQ SEQUENCE 110 AA; 11275 MW; D004935674B62BCE CRC64;

Query Match 31.8%; Score 46.5; DB 2; Length 110;
Best Local Similarity 46.4%; Pred. No. 31;
Matches 13; Conservative 2; Mismatches 10; Indels 3; Gaps 1;

QY 3 EGLYQGVPRAPG---TEARRHYDEGVR 27 | :|||:|
DB 75 EKVKAGVDRAKAGIHNAEHASYDEGHR 102 | :|||:|

RESULT 33
Q9N3R9 ID Q9N3R9 PRELIMINARY; PRT; 4158 AA.
AC Q9N3R9;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE Y47G6A.23 PROTEIN.
GN Y47G6A.23.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=BRISTOL N2;
RA Minx P.; Graves T.; Hawrysko C.;
RT "The sequence of C. elegans cosmid Y47G6A."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024791; AAF60650.1; -
SQ SEQUENCE 4158 AA; 469192 MW; 3A7A285E107F705D CRC64;

Query Match 31.8%; Score 46.5; DB 5; Length 4158;
Best Local Similarity 42.3%; Pred. No. 1.5e+03;
Matches 11; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

QY 3 EGLY---QGVPRAPGTEARRHYDEG 25 ||| :|||:|
DB 2482 DGLYQAEQGVSSOKPSTDKLIYREG 2507 ||| :|||:|

RESULT 34
Q9UE24 ID Q9UE24 PRELIMINARY; PRT; 341 AA.
AC Q9UE24;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE MRNA 3'-REGION (UNKNOWN FUNCTION) (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

Q9UZA5
ID Q9UZA5 PRELIMINARY; PRT; 592 AA.
AC Q9UZA5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE CARBON STARVATION PROTEIN A.
GN PAB1554.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248287; CAB50154.1; -
DR InterPro; IPR001202; -
DR PROSITE; PS01159; WW DOMAIN 1; UNKNOWN 1.
SQ SEQUENCE 592 AA; 64396 MW; E4B35BB9B6561638 CRC64;

Query Match 31.5%; Score 46; DB 1; Length 592;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 12; Conservative 2; Mismatches 8; Indels 2; Gaps 1;

QY 2 GEGLYQGVPRAPG--TEARRHYD 23
Db 22 GKGLQNKVVVRADPNRPTPAHRLYD 45

RESULT 39
Q9ULU3
ID Q9ULU3 PRELIMINARY; PRT; 618 AA.
AC Q9ULU3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE KIAA1126 PROTEIN (FRAGMENT).
GN KIAA1126.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA MEDLINE=20039618; PubMed=10574461;
RA Hirotsawa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT "Characterization of cDNA clones selected by the Genemark analysis from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
DR EMBL; AB032952; BAA86440.1; -
FT NON_TER 1
SQ SEQUENCE 618 AA; 67578 MW; 312E44EE31C0E5C4 CRC64;

Query Match 31.5%; Score 46; DB 4; Length 618;
Best Local Similarity 29.6%; Pred. No. 2.3e+02;
Matches 8; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

QY 1 VEGGLYQGVPRAPGCTEARRHYDEGR 27
Db 383 MQGVIFEGDPKAPSNSTAWQAYNAGVK 409

RESULT 40
Q9F584
ID Q9F584 PRELIMINARY; PRT; 774 AA.
AC Q9F584;
DT 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE RIORF171 PROTEIN.
GN RIORF171.
OS Agrobacterium rhizogenes.
OG Plasmid pRI1724.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N., Yoshida K.;
RT "The complete nucleotide sequence of a Ri (root inducing) plasmid indicates its chimerical structure between Ti and Sym plasmids.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid, pRI1724, by the construction of its physical map and library.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
RT "Genome structure of Ri plasmid (1): Construction of linking library and physical map of pRI1724 in Japanese Agrobacterium.";
RL Nucleic Acids Symp. Ser. 39:189-190(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N., Yoshida K.;
RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and its flanking regions of pRI1724 in Japanese Agrobacterium rhizogenes.";
RL Nucleic Acids Symp. Ser. 42:67-68(1999).
DR EMBL; AP002086; BAB16290.1; -
KW Plasmid.
SQ SEQUENCE 774 AA; 86002 MW; 8BBC516E30E8AD01 CRC64;

Query Match 31.5%; Score 46; DB 2; Length 774;
Best Local Similarity 45.2%; Pred. No. 3e+02;
Matches 14; Conservative 3; Mismatches 6; Indels 8; Gaps 2;

QY 1 VEGG-----LYQGVPRAPGCTEARRHYDEGV 26
Db 43 VGESTSTFVYGVPLAE---GERRAYEEW 70

RESULT 41
Q19206
ID Q19206 PRELIMINARY; PRT; 1741 AA.
AC Q19206;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CODED FOR BY C. ELEGANS CDNA CM21C7.
GN F08F8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,


```
ID Q9L6T0 PRELIMINARY; PRT; 309 AA.
AC Q9L6T0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE S. TYPHIMURIUM BRANCHED-CHAIN-AMINO-ACID TRANSAMINASE (ILVE)
DE (SP:P15168);
GN ILVE.
OS Salmonella typhimurium LT2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=99287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SGSC1412;
RA WashU;
RT "The Salmonella typhimurium Genome Sequencing Project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SGSC1412;
RA Waterston R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF233324; AAF33481.1; -.
DR InterPro; IPR001544; -.
DR Pfam; PF01063; aminotran_4; 1.
DR PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
SQ SEQUENCE 309 AA; 34052 MW; 4F60941F3D63CC69 CRC64;

Query Match 31.2%; Score 45.5; DB 2; Length 309;
Best Local Similarity 31.4%; Pred. No. 1.3e+02;
Matches 16; Conservative 3; Mismatches 5; Indels 27; Gaps 3;

QY 3 EGLYQGV-----RAEP-----GTEARRH-YDEGV 26
| | | | |
| | | | |
Db 134 EALDQGDAMVSSNRAAPNTIPTAAKAGNYLSLLVGSARRHG YQEG 184

RESULT 45
O66027 PRELIMINARY; PRT; 680 AA.
AC O66027;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE METAL TRANSPORTING ATPASE MTA72.
GN MTA7.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ERDMAN, ATCC 35801;
RX MEDLINE=98181048; PubMed=9514635;
RA Calder K.W., Horwitz M.A.;
RT "Identification of iron-regulated proteins of Mycobacterium tuberculosis and cloning of tandem genes encoding a low iron-induced protein and a metal transporting ATPase with similarities to two-component metal transport systems.";
RT Microb. Pathog. 24:133-143(1998).
RL EMBL; U82820; AAC15948.1; -.
DR InterPro; IPR000150; -.
DR InterPro; IPR001454; -.
DR InterPro; IPR001757; -.
DR Pfam; PF00122; EI-E2_ATPase; 1.
DR Pfam; PF00702; Hydrolase; 1.
DR PRINTS; PR00119; CATATPASE.
DR PROSITE; PS00154; ATPASE_EI_E2; UNKNOWN_1.
DR PROSITE; PS01229; COF_2; UNKNOWN_1.
SQ SEQUENCE 680 AA; 72149 MW; 0368D58CB53F30BD CRC64;
```

```
Query Match 31.2%; Score 45.5; DB 2; Length 680;
Best Local Similarity 44.0%; Pred. No. 3e+02;
Matches 11; Conservative 2; Mismatches 11; Indels 1; Gaps 1;

QY 2 GEGLYQGV-RAEPGTEARRHYDEG 25
| | | | |
| | | | |
Db 67 GTGLARGAPLVRRAATAARNHRADG 91
```

Search completed: June 28, 2001, 11:55:38
Job time: 159 sec

RESULT	1	
AAW71869		
ID	AAW71869	standard; Protein; 553 AA.
XX	XX	
XX	AAW71869;	
XX	XX	
DT	06-JAN-1999	(first entry)
XX	XX	
DE	Amino acid encoded by prostate tumour clone L1-12.	
XX	XX	
KW	XX	Prostate; cancer; tumour; vaccine; immunogen; clone
XX	XX	
OS	Homo sapiens.	
XX	XX	

ALIGNMENTS

Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers

Example 1; Page 87-89; 141pp; English.

This sequence is encoded by a human prostate tumour specific gene, and is used in the method of the invention. The method is for detecting prostate cancer comprising contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein, such as this protein sequence. An antibody which binds to the immunogenic portion of the prostate protein and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may also be conjugated to a therapeutic agent for use in therapy of prostate cancers.

Sequence 553 AA;

Query Match 100.0%; Score 88; DB 19; Length 553;
Best Local Similarity 100.0%; Pred. No. 8.6e-07; Indels 0; Gaps 0;
Matches 16; Conservative 0; Mismatches 0;

1 DWDTSALAPYLGTQEE 16
182 dwdtsalapylgtgee 197

RESULT 3
AAB28527
ID AAB28527 standard; Protein; 553 AA.

AA28527;
AC AAB28527;
DT 07-FEB-2001 (first entry)
DX
DX
DX
KW Human; breast tumour antigen; cytostatic; immunotherapy;
KW breast cancer; vaccine.
XX Homo sapiens.
XX
XX WO2000061756-A2.
PD 19-OCT-2000.
PF 10-APR-2000; 2000WO-US09688.
XX
XX 09-APR-1999; 99US-0288950.
PR 02-JUL-1999; 99US-0346327.
XX
XX (CORI-) CORIXA CORP.
PA
XX Reed SG, Xu J, Dillon DC;
XX WPI: 2000-638568/61.
XX N-PSDB; AAC79473.
XX
XX A novel isolated polypeptide comprising an immunogenic portion of a breast cancer protein useful in the detection and treatment of breast cancer -
PT
PT
XX Claim 2; Page 92-93; 95pp; English.
XX
XX The present sequence is encoded by a cDNA sequence which was isolated from a breast tumour cDNA library. It is provided as a specific antigen relating to compounds for immunotherapy, diagnosis of breast cancer. Breast tumour antigens and polynucleotides that encode them may be used in the treatment of a pharmaceutical composition to be used in the treatment of breast cancer. Proliferated T cells and incubated antigen presenting cells are also required. The polypeptides and polynucleotides may also be used to produce a vaccine.

SQ Sequence 553 AA;

Query Match 100.0%; Score 88; DB 21; Length 553;
 Best Local Similarity 100.0%; Pred. No. 8.6e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQEE 16
 |||||
 Db 182 dwdtsalapytgtqee 197

RESULT 4

AA82002
 ID AAY82002 standard; Protein; 553 AA.

XX
 AC AAY82002;

XX
 DT 13-JUN-2000 (first entry)

XX
 DE Human immunogenic prostate tumour protein sequence SEQ ID NO:113.

XX
 KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
 KW immunogenic; cytostatic; vaccine.

XX
 OS Homo sapiens.

XX
 PN WO200004149-A2.

XX
 PD 27-JAN-2000.

XX
 PF 14-JUL-1999; 99WO-US15838.

XX
 PR 14-JUL-1998; 98US-0115453.

XX
 PR 14-JUL-1998; 98US-0116134.

XX
 PR 23-SEP-1998; 98US-0159812.

XX
 PR 23-SEP-1998; 98US-0159822.

XX
 PR 15-JAN-1999; 99US-0232149.

XX
 PR 15-JAN-1999; 99US-0232880.

XX
 PR 09-APR-1999; 99US-0288946.

XX
 PA (CORI-) CORIXA CORP.

XX
 PI Dillon DC, Harlocker SL, Yuguu J, Xu J, Mitcham JL;

XX
 WPI; 2000-171268/15.

XX
 PT New polypeptide useful for treating and diagnosing prostate cancer

XX
 PT comprises an immunogenic portion of prostate tumor protein -

XX
 PS Claim 3; Page 138-139; 263pp; English.

XX
 CC The present invention describes isolated polypeptides, comprising an
 CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
 CC and polynucleotides encoding them have cytostatic activity and can be
 CC used in vaccines and in gene therapy. The polypeptides and
 CC polynucleotides encoding them, antigen presenting cells which express
 CC the polypeptides, antibodies against the polypeptides and vaccines
 CC comprising them can be used for inhibiting the development of prostate
 CC cancer in a patient. The polypeptides can be used to generate antibodies
 CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
 CC the polynucleotides encoding the polypeptides can be used as a probe or
 CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
 CC AAY82000 to AAY82020 represent sequences used in the exemplification of
 CC the present invention.

SQ Sequence 553 AA;

Query Match

Best Local Similarity 100.0%; Score 88; DB 21; Length 553;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQEE 16
 |||||
 Db 182 dwdtsalapytgtqee 197

RESULT 5

AA82002
 ID AAR97244 standard; Protein; 4473 AA.

XX
 AC AAR97244;

XX
 DT 07-JAN-1997 (first entry)

XX
 DE Virulence gene cluster polypeptide product.

XX
 KW Mutant; adaptation; virulence factor; identification; screening;
 KW vaccine; drugs; infection; treatment.

XX
 OS Salmonella typhimurium.

XX
 FH Key Location/Qualifiers

XX
 FT Region

XX
 FT /note= "All x's in this sequence correspond to
 FT termination codons in the virulence gene
 FT cluster sequence given in AAT09224."

XX
 PN WO9617951-A2.

XX
 PD 13-JUN-1996.

XX
 PF 11-DEC-1995; 95WO-GB02875.

XX
 PR 05-MAY-1995; 95GB-0009239.

XX
 PR 09-DEC-1994; 94GB-0024921.

XX
 PR 31-JAN-1995; 95GB-0001881.

XX
 PA (RPMS-) RPMS TECHNOLOGY LTD.

XX
 PI Holden DW;

XX
 WPI; 1996-287194/29.

XX
 N-PSDB; AAT09224.

XX
 PT Identifying virulence genes in microorganisms - by introducing

XX
 PT mutants with insertion inactivated genes into environment and

XX
 PT retrieval and analysis of mutants

XX
 PS Claim 51; Figure 11; 131pp; English.

XX
 CC A method for identifying a microorganism having a reduced adaptation
 CC to a particular environment comprising the steps of: (1) providing a
 CC plurality of microorganisms each of which is independently mutated by
 CC the insertional inactivation of a gene with a nucleic acid comprising
 CC a unique marker sequence so that each mutant contains a different
 CC marker sequence, or clones of the said microorganism; (2) providing
 CC individually a stored sample of each mutant produced by step (1) and
 CC providing individually stored nucleic acid comprising the unique
 CC marker sequence from each individual mutant; (3) introducing a
 CC plurality of mutants produced by step (1) into the said particular
 CC environment and allowing those microorganisms which are able to do so
 CC to grow in the said environment; (4) retrieving microorganisms from
 CC the said environment or a selected part thereof and isolating the
 CC nucleic acid from the retrieved microorganisms; (5) comparing any
 CC marker sequences in the nucleic acid isolated in step (4) to the
 CC unique marker sequence of each individual mutant stored as in step
 CC (2); and (6) selecting an individual mutant which does not contain any
 CC of the marker sequences as isolated in step (4). The products and
 CC methods can be used for identifying virulence genes in microorganisms.
 CC The mutant microorganisms can be used in vaccines or to screen for
 CC drugs which reduce virulence or compounds useful for preventing,
 CC ameliorating or treating infections in animals or plants.

XX
 SQ Sequence 4473 AA;

AAW18187
ID AAW18187 standard; peptide; 22 AA.

XX AC AAW18187;
XX DT 13-AUG-1997 (first entry)
XX DE N-terminal sequence from Mycobacterium tuberculosis 23KD protein.
XX DE Vaccine: vaccinating agent; M.tuberculosis; pathogen; bacteria;
XX KW virus; fungus; protozoan; HIV.

XX OS Mycobacterium tuberculosis.
XX PN W09637219-A1.

XX PD 28-NOV-1996.

XX PF 23-MAY-1996; 96WO-US07781.

XX PR 06-DEC-1995; 95US-0568357.

XX PR 23-MAY-1995; 95US-0447398.

XX PR 20-OCT-1995; 95US-0545926.

XX PR 31-OCT-1995; 95US-0551149.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Harth G, Horwitz MA;

XX DR WPI; 1997-020936/02.

XX PT Vaccines derived from M.tuberculosis major abundant extracellular
XX PT proteins - are easy to prepare and less toxic than conventional
XX PT killed or attenuated vaccines, useful for protecting against or
XX PT treating Mycobacterial infections

XX PS Example 2; Page 110; 193pp; English.

XX CC A vaccinating agent for promoting an immune response in a mammal
XX CC against Mycobacterium pathogens comprises at least one
XX CC immunodominant epitope of at least one majorly abundant
XX CC extracellular protein, i.e. the M.tuberculosis 110, 80, 71, 58, 45,
XX CC 32A, 32B, 30, 24, 23.5, 23, 16, 14 or 12 kD proteins, or their
XX CC analogues, homologues and subunits. The present sequence represents
XX CC the N-terminal amino acid sequence of the 23 kD protein. The N-terminal
XX CC amino acid sequences of the major abundant extracellular products were
XX CC determined to provide structural data and to uncover possible
XX CC relationships between the proteins. The vaccinating agents are used to
XX CC protect against (or to treat existing) infections by Mycobacterium
XX CC (especially M.tuberculosis) while the epitopes can also be used to
XX CC detect presence of an immune response to a Mycobacterium pathogen. The
XX CC vectors, containing the DNA for the extracellular proteins, are used to
XX CC transform cells for production of recombinant DNA molecules. More
XX CC generally the DNA from other pathogens can be used in vaccines, e.g.
XX CC against other bacteria, viruses, fungi and protozoa. Since different
XX CC combinations of DNA can be used, a wide range of effective compositions
XX CC can be produced. They generate a response against the antigens most
XX CC often found on infected cells during the infection, regardless of the
XX CC strength or specificity of the immune response. The vaccines are easy to
XX CC produce and less toxic than known killed or attenuated vaccines, so can
XX CC be given to immunocompromised subjects, e.g. those with HIV infection.

XX SQ Sequence 22 AA;

Query Match 47.7%; Score 42; DB 18; Length 22;
Best Local Similarity 50.0%; Pred. No. 1.6;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLGTQ 14

DB 9 dwdygalephisgg 22

RESULT 9

AAW75575
ID AAW75575 standard; protein; 22 AA.

XX AC AAW75575;

XX DT 23-OCT-1998 (first entry)

XX DE M. tuberculosis 23 kD extracellular product N-terminal sequence.
XX KW Mycobacterium tuberculosis; vaccination; extracellular product;
XX KW immunodominant epitope; interleukin-12; MF59; immune response;
XX KW opsonising humoral response; intracellular pathogen.

XX OS Mycobacterium tuberculosis.

XX PN W09831388-A1.

XX PD 23-JUL-1998.

XX PF 15-JAN-1998; 98WO-US00942.

XX PR 21-JAN-1997; 97US-0786533.

XX PA (REGC) UNIV CALIFORNIA.

XX PI Harth G, Horwitz MA, Lee B;

XX DR WPI; 1998-413815/35.

XX PT Vaccines against Mycobacterium containing major extracellular
XX PT proteins - used to, e.g. induce protective and therapeutic immune
XX PT responses, and for detecting an immune response

XX PS Example 2; Page 35; 236pp; English.

XX CC Sequences shown in AAW75570 to AAW75586 represent N-terminal amino acid
XX CC sequences of 14 exemplary major abundant extracellular products of
XX CC Mycobacterium tuberculosis. The invention provides an agent for
XX CC vaccinating mammals against Mycobacterium. The agent comprises at least
XX CC one of the major abundant extracellular 110, 80, 71, 58, 45, 32A, 32B,
XX CC 30, 24, 23.5, 23, 16, 14 or 12 kDa proteins of M. tuberculosis, or at
XX CC least 1 of their immunodominant epitopes and interleukin-12 (IL-12) or
XX CC MF59 as adjuvants. The agent containing the nucleic acid encoding the
XX CC extracellular products are used to raise a protective or therapeutic
XX CC immune response against Mycobacterium, specifically M. tuberculosis.
XX CC The immunodominant epitopes can also be used (typically in a cutaneous
XX CC hypersensitivity test) to detect an immune response to vaccination.
XX CC Preparation of the agent does not require selection of the most
XX CC immunogenic products, so large scale production and purification are
XX CC easy, resulting in a consistent, standardised formulation, having lower
XX CC toxicity than killed or attenuated vaccines. The agents provide a rapid
XX CC and effective response (including a strong cell-mediated component) and
XX CC are safe even in immunocompromised subjects. They prevent development of
XX CC an opsonising humoral response that might spread intracellular
XX CC pathogens.

XX SQ Sequence 22 AA;

Query Match 47.7%; Score 42; DB 19; Length 22;
Best Local Similarity 50.0%; Pred. No. 1.6;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLGTQ 14

DB 9 dwdygalephisgg 22

RESULT 10

AAW58854
ID AAW58854 standard; Protein; 121 AA.

Human gene 36 encoded secreted protein HLHDL42, SEQ ID NO:118.

Human; secreted protein; proliferative disorder; cancer; tumour;
foetal abnormality; developmental abnormality; haematopoietic disorder;
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
inflammation; allergy; neurological disorder; Alzheimer's disease;
Parkinson's disease; cognitive disorder; schizophrenia; asthma;
skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
cardiovascular disorder; angiogenic disorder; kidney disorder;
gastrointestinal disorder; pregnancy-related disorder;
endocrine disorder; infection; wound healing; vulneryary;
cell culture; chemotaxis; food additive;
binding partner identification.

Homo sapiens.

WO200118022-A1.

15-MAR-2001.

31-AUG-2000; 2000WO-US24008.

03-SEP-1999; 99US-0152315.

03-SEP-1999; 99US-0152317.

(HUMA-) HUMAN GENOME SCI INC.

Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
Moore PA, Shi Y, Wei Y, Florence KA;

WPI: 2001-203081/20.

N-PSDB; AAF91893.

Nucleic acid molecules encoding human secreted proteins, used in
preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
Parkinson's diseases and cancers -

Claim 11; Page 553; 607pp; English.

AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
protein genes, and AAB87342-AAB87413 represent the proteins they encode.
AAB87414-AAB87454 represent human secreted protein fragments. The genes
and their corresponding secreted proteins are useful for preventing,
treating or ameliorating medical conditions, e.g., by protein or gene
therapy. Pathological conditions can be diagnosed by determining the
amount of the new protein in a sample or by determining the presence of
mutations in the new genes. Specific uses are described for each of the
52 genes, based on the tissues in which they are most highly expressed,
and include developing products for the diagnosis or treatment of
proliferative disorders, cancer, tumours, foetal and developmental
abnormalities, haematopoietic disorders, diseases of the immune system,
AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
allergies, neurological disorders (e.g., Alzheimer's disease,
Parkinson's disease), cognitive disorders, schizophrenia, asthma,
skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
cardiovascular disorders, angiogenic disorders, kidney disorders,
gastrointestinal disorders, pregnancy-related disorders, endocrine
disorders, and infections. The proteins can also be used to aid wound
healing and epithelial cell proliferation, to prevent skin aging due to
sunburn, to maintain organs before transplantation, for supporting cell
culture of primary tissues, to regenerate tissues, to identify their
cognate ligands or binding partners, and in chemotaxis, and can be used
as a food additive or preservative to modify storage properties.
Antibodies specific for a protein of the invention can be used in
alleviating symptoms associated with the disorders mentioned above, and
in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
immunosorbent assay (ELISA). The present sequence represents a human
secreted protein of the invention.

Sequence 53 AA;

Query Match 46.6%; Score 41; DB 22; Length 53;
Best Local Similarity 72.7%; Pred. No. 6.1;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 SALAPYLGTQE 15
||| |||| :|
Db 41 salqpylgkee 51

RESULT 13
AAB87407
ID AAB87407 standard; Protein; 53 AA.
XX
AC AAB87407;
XX
DT 22-MAY-2001 (first entry)
XX
DE Human gene 36 encoded secreted protein HAPQU71, SEQ ID NO:148.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiogenic disorder; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulneryary;
KW cell culture; chemotaxis; food additive;
KW binding partner identification.
XX
OS Homo sapiens.
XX
PN WO200118022-A1.
XX
PD 15-MAR-2001.
XX
PF 31-AUG-2000; 2000WO-US24008.
XX
PR 03-SEP-1999; 99US-0152315.
XX
PR 03-SEP-1999; 99US-0152317.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
DR WPI: 2001-203081/20.
DR N-PSDB; AAF91923.
XX
PT Nucleic acid molecules encoding human secreted proteins, used in
preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
Parkinson's diseases and cancers -
XX
PS Claim 11; Page 574; 607pp; English.
XX
AAAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
protein genes, and AAB87342-AAB87413 represent the proteins they encode.
AAB87414-AAB87454 represent human secreted protein fragments. The genes
and their corresponding secreted proteins are useful for preventing,
treating or ameliorating medical conditions, e.g., by protein or gene
therapy. Pathological conditions can be diagnosed by determining the
amount of the new protein in a sample or by determining the presence of
mutations in the new genes. Specific uses are described for each of the
52 genes, based on the tissues in which they are most highly expressed,
and include developing products for the diagnosis or treatment of
proliferative disorders, cancer, tumours, foetal and developmental
abnormalities, haematopoietic disorders, diseases of the immune system,
AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
allergies, neurological disorders (e.g., Alzheimer's disease,
Parkinson's disease), cognitive disorders, schizophrenia, asthma,
skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
cardiovascular disorders, angiogenic disorders, kidney disorders,
gastrointestinal disorders, pregnancy-related disorders, endocrine
disorders, and infections. The proteins can also be used to aid wound
healing and epithelial cell proliferation, to prevent skin aging due to
sunburn, to maintain organs before transplantation, for supporting cell
culture of primary tissues, to regenerate tissues, to identify their
cognate ligands or binding partners, and in chemotaxis, and can be used
as a food additive or preservative to modify storage properties.
Antibodies specific for a protein of the invention can be used in
alleviating symptoms associated with the disorders mentioned above, and
in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
immunosorbent assay (ELISA). The present sequence represents a human
secreted protein of the invention.


```
XX WPI; 2000-638177/61.
DR N-PSDB; AAC59144.
XX
PT Novel nucleic acids encoding 49 human secreted proteins useful for
PT treating cancers, hyperproliferative disorders, inflammatory disorders,
PT neurological disorders and cardiovascular disorders
XX
XX Claim 11; Page 359; 389pp; English.
XX
XX Sequences AAB28012-B28060 represent the amino acid sequences of 49
XX human secreted proteins encoded by the genes AAC59108-C59156. The genes
XX and proteins are useful for preventing, ameliorating or treating medical
XX conditions, e.g. by protein or gene therapy. The genes are isolated from
XX a range of human tissues disclosed in the specification. The nucleic
XX acids, proteins, antibodies and (ant)agonists are useful in the
XX diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
XX ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
XX Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX
XX Sequence 54 AA;
SQ
Query Match 46.6%; Score 41; DB 21; Length 54;
Best Local Similarity 72.7%; Pred. No. 6.2;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Oy 5 SALAPYLGTQE 15
Db ||| |||| :|
41 salqpylgkee 51
RESULT 16
AAB75548
ID AAB75548 standard; Protein; 54 AA.
XX
AC AAB75548;
XX
DT 06-APR-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 43 SEQ ID NO:102.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neutropic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; vulnary; autoimmune disease; cardiovascular disorder;
KW hyperproliferative disorders; cerebrovascular disorder; wound healing;
KW nervous system disorder; ocular disorder; skin ageing; chemotaxis;
KW food additive.
XX
OS Homo sapiens.
XX
PN WO200077026-A1.
XX
PD 21-DEC-2000.
XX
PF 01-JUN-2000; 2000WO-US14973.
XX
PR 11-JUN-1999; 99US-0138630.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX WPI; 2001-071258/08.
DR
N-PSDB; AAF64218.
Nucleic acid molecules encoding human secreted proteins, used in
preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
Parkinson's diseases and cancers -
Claim 11; Page 489; 542pp; English.
Human secreted proteins AAB75506 - AAB75554 are encoded by polynucleotide
sequences AAF64176 - AAF64224. The specification includes amino acid
secreted proteins, and protein sequences with which they share homology.
The proteins and polynucleotides, their agonists and antagonists have
activities dependent on the tissues and cells in which they are
expressed, examples of these activities include, immunosuppressive;
antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
vasotropic; cerebroprotective; neutropic; neuroprotective; antibacterial;
virucide; fungicide; ophthalmological; and vulnary. The proteins,
polynucleotides, agonists and antagonists can be used to treat or detect
or diagnose various diseases and disorders including, autoimmune
diseases e.g. rheumatoid arthritis, hyperproliferative disorders
e.g. neoplasms of the breast or liver, cardiovascular disorders
e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
angiogenesis, nervous system disorders e.g. Alzheimer's disease,
infections caused by bacteria, viruses and fungi and ocular disorders
e.g. corneal infection. The polypeptides can also be used to aid wound
healing and epithelial cell proliferation, to prevent skin ageing due to
sunburn, to maintain organs before transplantation, for supporting cell
culture of primary tissues, to regenerate tissues and in chemotaxis. The
polypeptides can also be used as a food additive or preservative to
increase or decrease storage capabilities. Included in the invention are
polynucleotide sequences AAF64167 - AAF64175 and peptide AAB75505 which
are used in the isolation, identification and characterisation of the
proteins of the invention.
Sequence 54 AA;
SQ
Query Match 46.6%; Score 41; DB 22; Length 54;
Best Local Similarity 72.7%; Pred. No. 6.2;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Oy 5 SALAPYLGTQE 15
Db ||| |||| :|
41 salqpylgkee 51
RESULT 17
AAB41270
ID AAB41270 standard; Protein; 122 AA.
XX
AC AAB41270;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1034 polypeptide sequence SEQ ID NO:2068.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiprosiatric; antiparkinsonian; neutropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antinflamatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflamatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
OS
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U932236; AAC51228.1; -;
 DR EMBL; U932237; AAC51229.1; -;
 DR EMBL; U93237; AAC51230.1; -;
 DR MIM; 131100; -;
 DR MIM; 145000; -;
 KW Nuclear protein; Disease mutation; Alternative splicing; Polymorphism.
 FT VARSPLIC 149 153 MISSING (IN SHORT ISOFORM).
 FT VARIANT 12 12 P -> L (IN FMEN1).
 FT VARIANT 22 22 L -> R (IN FMEN1).
 FT VARIANT 26 26 E -> K (IN PARATHYROID ADENOMA).
 FT VARIANT 39 39 L -> W (IN FMEN1).
 FT VARIANT 42 42 G -> D (IN FMEN1).
 FT VARIANT 45 45 E -> G (IN FMEN1).
 FT VARIANT 119 119 MISSING (IN FMEN1).
 FT VARIANT 135 135 K -> I (IN FMEN1).
 FT VARIANT 139 139 H -> D (IN FMEN1).
 FT VARIANT 139 139 H -> Y (IN FMEN1 AND SPORADIC MEN1).
 FT VARIANT 144 144 F -> V (IN FMEN1).
 FT VARIANT 161 161 G -> D (IN FMEN1).
 FT VARIANT 165 165 A -> P (IN FMEN1).
 FT VARIANT 169 169 A -> D (IN FMEN1).
 FT VARIANT 171 173 MISSING (IN FMEN1).
 FT VARIANT 176 176 /FTIG-Var_005439.
 FT R -> Q.
 Query Match 31.5%; Score 46; DB 1; Length 615;
 Best Local Similarity 52.9%; Pred. No. 47;
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
 QY 2 GEGLYQGVPRAPGTEA 18
 DB 512 GQGVSGPPKPPGTVA 528
 RESULT 13
 ID P73_CERAE STANDARD; PRT; 637 AA.
 AC Q9XS88; Q9TS09;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE TUMOR PROTEIN P73 (P53-LIKE TRANSCRIPTION FACTOR) (P53-RELATED
 DE PROTEIN).
 GN TP73 OR P73.
 OS Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
 OC Cercopithecinae; Cercopithecus.
 OX NCBI_TaxID=95534;
 RT [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;

RA Caput D.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PARTICIPATES IN THE APOPTOTIC RESPONSE TO DNA DAMAGE.
 CC WHEN OVERPRODUCED, ACTIVATES TRANSCRIPTION FROM P53-RESPONSIVE
 CC PROMOTERS AND INDUCES APOPTOSIS. MAY BE A TUMOR SUPPRESSOR
 CC PROTEIN (BY SIMILARITY).
 CC -!- SUBUNIT: THE C-TERMINAL OLIGOMERIZATION DOMAIN BINDS TO THE ABL
 CC TYROSINE KINASE SH3 DOMAIN. ISOFORM BETA INTERACTS HOMOTYPICALLY
 CC AND WITH P53, WHEREAS ISOFORM ALPHA DOES NOT.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ALPHA (SHOWN HERE) AND BETA;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- DOMAIN: POSSESSES AN ACIDIC TRANSACTIVATION DOMAIN, A CENTRAL DNA
 CC BINDING DOMAIN AND A C-TERMINAL OLIGOMERIZATION DOMAIN THAT BINDS
 CC TO THE ABL TYROSINE KINASE SH3 DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE P53 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y11419; CAA72224.1; -;
 DR EMBL; Y11419; CAA72225.1; -;
 DR HSP; P04637; LYCS
 DR InterPro; IPR002117; -;
 DR Pfam; PF00870; P53; 1.
 DR PRINTS; PR00386; P53SUPPRESSR.
 KW Transcription regulation; Activator; DNA-binding; Anti-oncogene;
 KW Apoptosis; Nuclear protein; Phosphorylation; Alternative splicing.
 FT VARSPLIC 495 637
 FT ELQQRVMEAVHVRVRRHTITIPNRGGPGAGPDEWADFGL
 FT PDCKARKQPIKEEFTAEATH -> RTWGP (IN ISOFORM
 FT BETA).
 SQ SEQUENCE 637 AA; 69630 MW; 7CB200B919C9C70A CRC64;
 Query Match 31.5%; Score 46; DB 1; Length 637;
 Best Local Similarity 56.2%; Pred. No. 48;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 QY 9 VPRAEPGTEARRHYDE 24
 DB 337 VPALGPGVKRRHGE 352
 RESULT 14
 ID IE63_HCMVA STANDARD; PRT; 744 AA.
 AC P16749;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE TRANSCRIPTIONAL REGULATOR IE63 HOMOLOG (PROTEIN UL69).
 GN UL69.
 OS Human cytomegalovirus (strain AD169).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 OX NCBI_TaxID=10360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90269039; PubMed=2161319;
 RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
 RA Horsnell T., Hutchison C.A. III, Kourzarides T., Martignetti J.A.,
 RA Freddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
 RT "Analysis of the protein-coding content of the sequence of human
 RT cytomegalovirus strain AD169."
 RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
 CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL54,

FT Misc-difference 117
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 184
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 277
 FT /label= substitution
 FT /note= "Lys to Arg substitution"
 FT Misc-difference 448
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 503..524
 FT /label= deletion
 FT /note= "amino acids 503 to 524 are opt. deleted"
 XX US5501853-A.
 PN 26-MAR-1996.
 PD 23-DEC-1985; 85US-0812879.
 PF 22-DEC-1986; 86US-0944117.
 PR 23-DEC-1985; 85US-0812879.
 XX (BEHW) BEHRINGER AG.
 PA (CHIR) CHIRON CORP.
 XX Afting E, Haigwood NL, Mullenbach G, Paques EP;
 PI WPI; 1996-179247/18.
 DR Human tissue plasminogen activator mutants for fibrin clot lysis -
 PT with improved plasminogenolytic specific activity, increased fibrin
 PT dependence and decreased plasminogen inhibitor susceptibility
 XX Claim 4; Page -: 8pp; English.
 PS AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants
 CC that have improved plasminogen specific activity, increased fibrin
 CC dependence and decreased plasminogen inhibitor susceptibility. The
 CC mutants are useful in the lysis of fibrin clots and in the
 CC prevention of blood clot formation by activating plasminogen. The
 CC specific activity of tPA is enhanced by reducing the amount of
 CC glycosylation of the protein by modifying the glycosylation sites at
 CC positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to
 CC plasminogen activator inhibition is reduced by modifying the cleavage
 CC site at position 274-278 and the fibrin dependence of tPA is increased
 CC by truncating the C-terminus.
 XX Sequence 524 AA;
 SQ Query Match 44.3%; Score 39; DB 17; Length 524;
 Best Local Similarity 44.4%; Pred. No. 1.7e+02;
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
 QY 1 DWDTSALA--PYLGTOEE 16
 Db 115 nwessalagkpsgrpd 132
 :|::||| | | : :
 RESULT 20
 AAR99133
 ID AAR99133 standard; protein: 524 AA.
 XX AAR99133;
 AC AAR99133;
 XX 29-OCT-1996 (first entry)
 DT Mutant human tissue plasminogen activator for fibrin clot lysis.
 DE Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
 XX

KW improved specificity; glycosylation sites; truncated C-terminus;
 KW deletion; substitution; fibrinogen.
 XX Synthetic.
 OS Key Location/Qualifiers
 FH Misc-difference 117
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 184
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 277
 FT /label= substitution
 FT /note= "Lys to Arg substitution"
 FT Misc-difference 503..524
 FT /label= deletion
 FT /note= "amino acids 503 to 524 are opt. deleted"
 XX US5501853-A.
 PN 26-MAR-1996.
 PD 23-DEC-1985; 85US-0812879.
 PF 22-DEC-1986; 86US-0944117.
 PR 23-DEC-1985; 85US-0812879.
 XX (BEHW) BEHRINGER AG.
 PA (CHIR) CHIRON CORP.
 XX Afting E, Haigwood NL, Mullenbach G, Paques EP;
 PI WPI; 1996-179247/18.
 DR Human tissue plasminogen activator mutants for fibrin clot lysis -
 PT with improved plasminogenolytic specific activity, increased fibrin
 PT dependence and decreased plasminogen inhibitor susceptibility
 XX Claim 4; Page -: 8pp; English.
 PS AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants
 CC that have improved plasminogen specific activity, increased fibrin
 CC dependence and decreased plasminogen inhibitor susceptibility. The
 CC mutants are useful in the lysis of fibrin clots and in the
 CC prevention of blood clot formation by activating plasminogen. The
 CC specific activity of tPA is enhanced by reducing the amount of
 CC glycosylation of the protein by modifying the glycosylation sites at
 CC positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to
 CC plasminogen activator inhibition is reduced by modifying the cleavage
 CC site at position 274-278 and the fibrin dependence of tPA is increased
 CC by truncating the C-terminus.
 XX Sequence 524 AA;
 SQ Query Match 44.3%; Score 39; DB 17; Length 524;
 Best Local Similarity 44.4%; Pred. No. 1.7e+02;
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
 QY 1 DWDTSALA--PYLGTOEE 16
 Db 115 nwessalagkpsgrpd 132
 :|::||| | | : :
 RESULT 21
 AAR99134
 ID AAR99134 standard; protein: 524 AA.
 XX AAR99134;
 AC AAR99134;
 XX 29-OCT-1996 (first entry)
 DT Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
 XX

XX 29-OCT-1996 (first entry)
 XX Mutant human tissue plasminogen activator for fibrin clot lysis.
 DE Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
 KW improved specificity; glycosylation sites; truncated C-terminus;
 KW deletion; substitution; fibrinogen.
 XX Synthetic.
 OS
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 117
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 184
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 277
 FT /label= substitution
 FT /note= "Lys to Arg substitution"
 FT Misc-difference 448
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 525
 FT /label= deletion
 FT /note= "Met-525 opt. deleted"
 FT Misc-difference 527
 FT /label= deletion
 FT /note= "Pro-527 opt. deleted"
 XX US5501853-A.
 XX 26-MAR-1996.
 XX 23-DEC-1985; 85US-0812879.
 XX 22-DEC-1986; 86US-0944117.
 XX 23-DEC-1985; 85US-0812879.
 XX (BEHW) BEHRINGERWERKE AG.
 XX (CHIR) CHIRON CORP.
 XX Afting E, Haigwood NL, Mullenbach G, Paques EP;
 XX WPI; 1996-179247/18.
 XX Human tissue plasminogen activator mutants for fibrin clot lysis -
 PT with improved plasminogenolytic specific activity, increased fibrin
 PT dependence and decreased plasminogen inhibitor susceptibility
 PS Claim 3; Page -: 8pp; English.
 XX AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants
 CC that have improved plasminogen specific activity, increased fibrin
 CC dependence and decreased plasminogen inhibitor susceptibility. The
 CC mutants are useful in the lysis of fibrin clots and in the
 CC prevention of blood clot formation by activating plasminogen. The
 CC specific activity of tPA is enhanced by reducing the amount of
 CC glycosylation of the protein by modifying the glycosylation sites at
 CC positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to
 CC plasminogen activator inhibition is reduced by modifying the cleavage
 CC site at position 274-278 and the fibrin dependence of tPA is increased
 CC by truncating the C-terminus.
 XX Sequence 527 AA;
 SQ
 Query Match 44.3%; Score 39; DB 17; Length 527;
 Best Local Similarity 44.4%; Pred. No. 1.7e+02;
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
 QY 1 DWDTSALA--PYLGTOEE 16

Db 115 nwessalaqkpysgrrpd 132
 :|::| | | | | | | | :
 RESULT 24
 AAR99125
 ID AAR99125 standard; protein; 527 AA.
 XX AAR99125;
 AC
 XX 29-OCT-1996 (first entry)
 DT
 XX Mutant human tissue plasminogen activator for fibrin clot lysis.
 DE
 XX Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
 KW improved specificity; glycosylation sites; truncated C-terminus;
 KW deletion; substitution; fibrinogen.
 XX
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 117
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 184
 FT /label= substitution
 FT /note= "Asn to Glu substitution"
 FT Misc-difference 277
 FT /label= substitution
 FT /note= "Lys to Arg substitution"
 FT Misc-difference 525
 FT /label= deletion
 FT /note= "Met-525 opt. deleted"
 FT Misc-difference 527
 FT /label= deletion
 FT /note= "Pro-527 opt. deleted"
 XX US5501853-A.
 XX 26-MAR-1996.
 XX 23-DEC-1985; 85US-0812879.
 XX 22-DEC-1986; 86US-0944117.
 XX 23-DEC-1985; 85US-0812879.
 XX (BEHW) BEHRINGERWERKE AG.
 XX (CHIR) CHIRON CORP.
 XX Afting E, Haigwood NL, Mullenbach G, Paques EP;
 XX WPI; 1996-179247/18.
 XX Human tissue plasminogen activator mutants for fibrin clot lysis -
 PT with improved plasminogenolytic specific activity, increased fibrin
 PT dependence and decreased plasminogen inhibitor susceptibility
 PS Claim 3; Page -: 8pp; English.
 XX AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants
 CC that have improved plasminogen specific activity, increased fibrin
 CC dependence and decreased plasminogen inhibitor susceptibility. The
 CC mutants are useful in the lysis of fibrin clots and in the
 CC prevention of blood clot formation by activating plasminogen. The
 CC specific activity of tPA is enhanced by reducing the amount of
 CC glycosylation of the protein by modifying the glycosylation sites at
 CC positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to
 CC plasminogen activator inhibition is reduced by modifying the cleavage
 CC site at position 274-278 and the fibrin dependence of tPA is increased
 CC by truncating the C-terminus.
 XX Sequence 527 AA;
 SQ

Query Match 44.3%; Score 39; DB 17; Length 527;
 Best Local Similarity 44.4%; Pred. No. 1.7e+02;
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

OY 1 DWDTSALA--PYLGTQEE 16
 DB 115 nwessalackpysgrprd 132

RESULT 26
 AAR99126
 ID AAR99126 standard; protein: 527 AA.
 AC AAR99126;
 XX 29-OCT-1996 (first entry)
 DE Mutant human tissue plasminogen activator for fibrin clot lysis.
 XX Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
 KW improved specificity; glycosylation sites; truncated C-terminus;
 KW deletion; substitution; fibrinogen.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 117 /label= substitution
 FT Misc-difference 277 /note= "Asn to Glu substitution"
 FT Misc-difference 448 /label= substitution
 FT Misc-difference 448 /note= "Lys to Arg substitution"
 FT Misc-difference 525 /label= substitution
 FT Misc-difference 525 /note= "Asn to Glu substitution"
 FT Misc-difference 527 /label= deletion
 FT Misc-difference 527 /note= "Met-525 opt. deleted"
 FT Misc-difference 527 /label= deletion
 FT Misc-difference 527 /note= "Pro-527 opt. deleted"
 XX US501853-A.
 XX 26-MAR-1996.
 XX 23-DEC-1985; 85US-0812879.
 XX 22-DEC-1986; 86US-0944117.
 XX 23-DEC-1985; 85US-0812879.
 XX (BEHW) BEHRINGERWERKE AG.
 XX (CHIR) CHIRON CORP.
 XX Afting E, Haigwood NL, Mullenbach G, Paques EP;
 XX WPI; 1996-179247/18.
 XX Human tissue plasminogen activator mutants for fibrin clot lysis -
 XX with improved plasminogenolytic specific activity, increased fibrin
 XX dependence and decreased plasminogen inhibitor susceptibility
 XX Claim 3; Page -; 8pp; English.
 XX AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants
 XX that have improved plasminogenolytic specific activity, increased fibrin
 XX dependence and decreased plasminogen inhibitor susceptibility. The
 XX mutants are useful in the lysis of fibrin clots and in the
 XX prevention of blood clot formation by activating plasminogen. The
 XX specific activity of tPA is enhanced by reducing the amount of
 XX glycosylation of the protein by modifying the glycosylation sites at
 XX positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to

CC plasminogen activator inhibition is reduced by modifying the cleavage
 CC site at position 274-278 and the fibrin dependence of tPA is increased
 CC by truncating the C-terminus.
 SQ Sequence 527 AA;

Query Match 44.3%; Score 39; DB 17; Length 527;
 Best Local Similarity 44.4%; Pred. No. 1.7e+02;
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

OY 1 DWDTSALA--PYLGTQEE 16
 DB 115 nwessalackpysgrprd 132

RESULT 26
 AAR99128
 ID AAR99128 standard; protein: 527 AA.
 AC AAR99128;
 XX 29-OCT-1996 (first entry)
 DE Mutant human tissue plasminogen activator for fibrin clot lysis.
 XX Tissue plasminogen activator; tPA; blood clot; coagulation; fibrin;
 KW improved specificity; glycosylation sites; truncated C-terminus;
 KW deletion; substitution; fibrinogen.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 117 /label= substitution
 FT Misc-difference 277 /note= "Asn to Glu substitution"
 FT Misc-difference 277 /label= substitution
 FT Misc-difference 525 /note= "Lys to Arg substitution"
 FT Misc-difference 525 /label= deletion
 FT Misc-difference 527 /note= "Met-525 opt. deleted"
 FT Misc-difference 527 /label= deletion
 FT Misc-difference 527 /note= "Pro-527 opt. deleted"
 XX US501853-A.
 XX 26-MAR-1996.
 XX 23-DEC-1985; 85US-0812879.
 XX 22-DEC-1986; 86US-0944117.
 XX 23-DEC-1985; 85US-0812879.
 XX (BEHW) BEHRINGERWERKE AG.
 XX (CHIR) CHIRON CORP.
 XX Afting E, Haigwood NL, Mullenbach G, Paques EP;
 XX WPI; 1996-179247/18.
 XX Human tissue plasminogen activator mutants for fibrin clot lysis -
 XX with improved plasminogenolytic specific activity, increased fibrin
 XX dependence and decreased plasminogen inhibitor susceptibility
 XX Claim 3; Page -; 8pp; English.
 XX AAR99124-R99138 are human tissue plasminogen activator (tPA) mutants
 XX that have improved plasminogenolytic specific activity, increased fibrin
 XX dependence and decreased plasminogen inhibitor susceptibility. The
 XX mutants are useful in the lysis of fibrin clots and in the
 XX prevention of blood clot formation by activating plasminogen. The
 XX specific activity of tPA is enhanced by reducing the amount of
 XX glycosylation of the protein by modifying the glycosylation sites at
 XX positions 117-119, 184-186 and/or 448-450. The sensitivity of tPA to

CC haemorrhage, heart stroke and pulmonary embolism.
CC See AAQ12071-2, AAR13150-53 and AAR12366-67 and also J03061482-84.
XX
SQ Sequence 558 AA;

Query Match 44.3%; Score 39; DB 12; Length 558;
Best Local Similarity 44.4%; Pred. No. 1.8e+02;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
QY 1 DWDTSALA--PYLGTOEE 16
DB 146 nwnssalaqepysgeepd 163

RESULT 29
AAR13151 standard; Protein; 559 AA.
XX AAR13151;
XX 29-AUG-1991 (first entry)
XX T-PA with -ve charged finger and/or kringle domain (4).
XX T-PA; tissue plasminogen activator; finger; kringle domain;
KW thrombosis; lysis.
XX Key Location/Qualifiers
FT Region 33..36 /label= R1
FT Region 37..41 /label= A2-6
FT Region 42..64 /note= "Tyr2-Cys6 of natural t-PA"
FT Region 65..157 /label= R2
FT Region 158..167 /label= A31-123
FT Region 168..169 /note= "Val31-Gln123 of natural t-PA"
FT Region 170..561 /label= R3
FT Region 170..561 /label= A136-527
FT Region 170..561 /note= "Leu136-Pro527 of natural t-PA"
XX JPO3061483-A.
XX 18-MAR-1991.
XX 26-JUN-1989; 89JP-0163599.
XX 27-JUN-1988; 88GB-0015246.
XX (FUJI) FUJISAWA PHARM KK.
XX WPI; 1991-180545/25.
XX N-PSDB; AAQ12071.
XX New tissue plasminogen activator with negatively charged finger
XX and/or kringle domain - is highly active and is useful as thrombosis-
XX lysing agent
XX Claim 1; Page 1; 29pp; Japanese.
XX This sequence is an example of a generic formula. The activator
XX is highly active in converting plasminogen to plasmin. It is
XX used as a thrombosis-lysing agent and is used in a pharmaceutical
XX company to treat diseases, such as cardiac infarction, cerebral
XX haemorrhage, heart stroke and pulmonary embolism.
XX See AAQ12071-2, AAR13150-53 and AAR12366-67 and also J03061482-84.
XX Sequence 559 AA;

Query Match 44.3%; Score 39; DB 12; Length 558;
Best Local Similarity 44.4%; Pred. No. 1.8e+02;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
QY 1 DWDTSALA--PYLGTOEE 16
DB 146 nwnssalaqepysgeepd 163

RESULT 30
AAR12367 standard; Protein; 561 AA.
XX AAR12367;
XX 29-AUG-1991 (first entry)
XX T-PA with -ve charged finger and/or kringle domain (7).
XX T-PA; tissue plasminogen activator; finger; kringle domain;
KW thrombosis; lysis.
XX Key Location/Qualifiers
FT Region 33..36 /label= R1
FT Region 37..41 /label= A2-6
FT Region 42..64 /note= "Tyr2-Cys6 of natural t-PA"
FT Region 65..157 /label= R2
FT Region 158..167 /label= A31-123
FT Region 168..169 /note= "Val31-Gln123 of natural t-PA"
FT Region 170..561 /label= R3
FT Region 170..561 /label= A136-527
FT Region 170..561 /note= "Leu136-Pro527 of natural t-PA"
XX JPO3061483-A.
XX 18-MAR-1991.
XX 26-JUN-1989; 89JP-0163599.
XX 27-JUN-1988; 88GB-0015246.
XX (FUJI) FUJISAWA PHARM KK.
XX WPI; 1991-180545/25.
XX N-PSDB; AAQ12071.
XX New tissue plasminogen activator with negatively charged finger
XX and/or kringle domain - is highly active and is useful as thrombosis-
XX lysing agent
XX Claim 1; Page 1; 29pp; Japanese.
XX This sequence is an example of a generic formula. The activator
XX is highly active in converting plasminogen to plasmin. It is
XX used as a thrombosis-lysing agent and is used in a pharmaceutical
XX company to treat diseases, such as cardiac infarction, cerebral
XX haemorrhage, heart stroke and pulmonary embolism.
XX See AAQ12071-2, AAR13150-53 and AAR12366-67 and also J03061482-84.
XX Sequence 561 AA;

Query Match 44.3%; Score 39; DB 12; Length 561;
Best Local Similarity 44.4%; Pred. No. 1.8e+02;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

OY 1 DWDTSALA--PYLGTQEE 16
:|:||||| || | :
Db 149 nwnssalapeysgeepd 166

RESULT 31
AARI2343
ID AARI2343 standard; Protein; 562 AA.
XX AC AARI2343;
XX DT 29-AUG-1991 (first entry)
XX DE T-PA with -ve charged finger and/or kringle domain (2).
XX KW T-PA; tissue plasminogen activator; finger; kringle domain;
XX KW thrombosis; lysis.
XX FH Key Location/Qualifiers
XX FT Region 33..36
XX FT /label= R1
XX FT Region 37..41
XX FT /label= A2-6
XX FT /note= "Tyr2-Cys6 of natural t-PA"
XX FT Region 42..65
XX FT /label= R2
XX FT Region 66..158
XX FT /label= A31-123
XX FT /note= "Val31-Gln123 of natural t-PA"
XX FT Region 159..170
XX FT /label= R3
XX FT Region 171..562
XX FT /label= A136-527
XX FT /note= "Leu136-Pro527 of natural t-PA"
XX PN JP03061483-A.
XX PD 18-MAR-1991.
XX PF 26-JUN-1989; 89JP-0163599.
XX PR 27-JUN-1988; 88GB-0015246.
XX PA (FUJI) FUJISAWA PHARM KK.
XX DR WPI; 1991-180545/25.
XX DR N-PSDB; AAQ12071.
XX PT New tissue plasminogen activator with negatively charged finger
XX PT and/or kringle domain - is highly active and is useful as thrombosis-
XX PT lysing agent
XX PS Disclosure; Fig 28; 29pp; Japanese.
XX CC This sequence is an example of a generic formula. The activator
XX CC is highly active in converting plasminogen to plasmin. It is
XX CC useful as a thrombosis-lysing agent and is used in a pharmaceutical
XX CC compsn. to treat diseases, such as cardiac infarction, cerebral
XX CC haemorrhage, heart stroke and pulmonary embolism.
XX CC See AAQ12071-2, AARI3150-53 and AARI2366-67 and also J03061482-84.
XX SQ Sequence 562 AA;

Query Match 44.3%; Score 39; DB 12; Length 562;
Best Local Similarity 44.4%; Pred. No. 1.8e+02;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

OY 1 DWDTSALA--PYLGTQEE 16
:|:||||| || | :
Db 150 nwnssalapeysgeepd 167

RESULT 32
AAR20210
ID AAR20210 standard; Protein; 734 AA.
XX AC AAR20210;
XX DT 16-APR-1992 (first entry)
XX DE Diacylglycerol kinase.
XX KW DGK; thymus cell; probe.
XX OS Sus scrofa.
XX PN JP03272689-A.
XX PD 04-DEC-1991.
XX PF 20-MAR-1990; 90JP-0071155.
XX PR 20-MAR-1990; 90JP-0071155.
XX PA (MEIP) MEIJI MILK PRODS KK.
XX DR WPI; 1992-027349/04.
XX DR N-PSDB; AAQ20670.
XX PT Diacylglycerol kinase which allows encoded cDNA to be isolated
XX PT - in which DNA contains base sequence which encodes amino acid
XX PT sequence of diacylglycerol kinase activity
XX PS Claim 4; Page 1 and Fig 1; 13pp; Japanese.
XX CC Total mRNA was extracted from swine thymus cells. The cDNA was
XX CC synthesised using oligo(dT) as primer and commercial BRL,
XX CC fractionated by agarose gel filtration to settled size
XX CC (e.g. at least 1 kb) and collected. A cDNA library was constructed
XX CC in pBR322 by using dc-dg homopolymer tailing method.
XX CC The DGK-encoding clones were isolated following hybridisation with
XX CC a probe. The sequence of each fragment was determined to obtain the
XX CC complete nucleotide sequence.
XX CC See AAQ21834-35 for probes.
XX SQ Sequence 734 AA;

Query Match 44.3%; Score 39; DB 13; Length 734;
Best Local Similarity 63.6%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 DWDTSALAPYL 11
||| | | | |
Db 149 dwdvselrpil 159

RESULT 33
AAG00538
ID AAG00538 standard; Protein; 65 AA.
XX AC AAG00538;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein, SEQ ID NO: 4619.
XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX KW gene therapy; chromosome mapping.
XX OS Homo sapiens.
XX PN EP1033401-A2.
XX PD 06-SEP-2000.

XX Prodn. of human monoclonal anti-tumour antibodies - by screening a
PT fusion phage library produced using peripheral blood lymphocytes
PT from a cancer patient
XX
XX Claim 19; Page 54; 82pp; English.
XX
XX A process for isolating and synthesising human monoclonal anti-tumour
CC antibodies has been produced. The process involves: (a) constructing at
CC least one fusion phage library from the peripheral blood lymphocytes
CC (PBLs) of a cancer patient; (b) screening for anti-tumour antibodies in
CC the phage library in a binding assay with cultured tumour cells of the
CC same type as the patient's tumour; (c) removing extraneous antibodies by
CC absorption against normal human cells; (d) cloning the phage selected in
CC step (b) and (c); (e) assaying the specificity of the cloned phage by
CC incubating the phage with at least two types of cultured normal cells;
CC and (f) further testing the specificity of cloned phage that do not bind
CC to either cell line of cultured normal cells in further binding assays
CC to cultured tumour cells derived from more than one other tumour that is
CC not the patient's tumour. The present sequence represents a human heavy
CC chain antibody, from an scFv antibody fusion phage library, produced by
CC a method as described above. The antibodies produced can be used for
CC diagnostic and therapeutic applications and for isolating tumour
CC antigens for studying tumourigenesis or for use as anti-cancer vaccines.
CC The human antibodies have low immunogenicity in humans compared to
CC murine monoclonal antibodies (MAbs). Since the antibodies are isolated
CC from fusion phage libraries, their affinity and specificity for a
CC tumour cell line can be improved by genetic manipulations.
XX
XX Sequence 118 AA;
SQ

Query Match 43.2%; Score 38; DB 18; Length 118;
Best Local Similarity 54.5%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 2 WDTSLAPYL 12
: : : : :
Db 71 wdtsixtaymg 81

RESULT 36
AAW04249
ID AAW04249 standard; Protein; 209 AA.
XX
XX AAW04249;
XX
XX 30-DEC-1996 (first entry)
XX
XX 4070A retrovirus amphotropic gp70 protein.
XX
XX Moloney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;
XX 10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;
XX polytropic MX27 provirus; targetted drug delivery; gene therapy;
XX single chain antibody; envelope protein.
XX
XX 4070A retrovirus.
XX
XX WO9630504-A1.
XX
XX 03-OCT-1996.
XX
XX 22-MAR-1996; 96WO-US03908.
XX
XX 24-MAR-1995; 95US-0409648.
XX
XX (GENE-) GENETIC THERAPY INC.
XX (UYSC-) UNIV SOUTHERN CALIFORNIA.
XX
XX Anderson W, Chiang YL, Januszski M, Mackrell AJ;
XX Zhao Y;
XX
XX WPI; 1996-455352/45.
XX
XX 03-OCT-1996.
XX
XX 22-MAR-1996; 96WO-US03908.
XX
XX 24-MAR-1995; 95US-0409648.
XX
XX (GENE-) GENETIC THERAPY INC.
XX (UYSC-) UNIV SOUTHERN CALIFORNIA.
XX
XX Anderson W, Chiang YL, Januszski M, Mackrell AJ;
XX Zhao Y;
XX
XX WPI; 1996-455352/45.

DR N-PSDB; AAT33923.
XX
XX Cell-targetted retroviral vector particles - having envelope protein
PT modified with targetting polypeptide
XX
XX Claim 1; Page 44-45; 73pp; English.
XX
XX Cell targetted retroviral vector particles can be used in gene
CC therapy to deliver a heterologous gene to a target cell for
CC expression of a heterologous polypeptide in that cell. The cell
CC targetted retroviral vector particles comprise an envelope protein
CC which is modified to contain a targetting polypeptide (a single chain
CC antibody). The preferred modifications to this sequence comprise the
CC replacement of amino acids 47-75, 47-93, 37-163.
XX
XX Sequence 209 AA;
SQ

Query Match 43.2%; Score 38; DB 17; Length 209;
Best Local Similarity 50.0%; Pred. No. 91;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 1 DWDTSALAPYL 12
: : : : :
Db 53 ewdpsdqepvg 64

RESULT 37
AAW04250
ID AAW04250 standard; Protein; 209 AA.
XX
XX AAW04250;
XX
XX 30-DEC-1996 (first entry)
XX
XX 10A1 murine leukaemia virus envelope protein.
XX
XX Moloney murine leukaemia virus; gp70; 4070A retrovirus; retrovirus;
XX 10A1 murine leukaemia virus; NZB-9-1 murine leukaemia virus;
XX polytropic MX27 provirus; targetted drug delivery; gene therapy;
XX single chain antibody; envelope protein.
XX
XX 10A1 murine leukaemia virus.
XX
XX WO9630504-A1.
XX
XX 03-OCT-1996.
XX
XX 22-MAR-1996; 96WO-US03908.
XX
XX 24-MAR-1995; 95US-0409648.
XX
XX (GENE-) GENETIC THERAPY INC.
XX (UYSC-) UNIV SOUTHERN CALIFORNIA.
XX
XX Anderson W, Chiang YL, Januszski M, Mackrell AJ;
XX Zhao Y;
XX
XX WPI; 1996-455352/45.
XX
XX N-PSDB; AAT33924.
XX
XX Cell-targetted retroviral vector particles - having envelope protein
PT modified with targetting polypeptide
XX
XX Claim 1; Page 45-47; 73pp; English.
XX
XX Cell targetted retroviral vector particles can be used in gene
CC therapy to deliver a heterologous gene to a target cell for
CC expression of a heterologous polypeptide in that cell. The cell
CC targetted retroviral vector particles comprise an envelope protein
CC which is modified to contain a targetting polypeptide (a single chain
CC antibody). The preferred modifications to this sequence comprise the
CC replacement of amino acids 47-75, 47-93, 47-163.

CC of (II) or for treatment of cancer pathologies (all claimed).

XX Sequence 209 AA;

Query Match 43.2%; Score 38; DB 17; Length 209;
Best Local Similarity 50.0%; Pred. No. 91;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLG 12
Db 53 ewdpsdqepvvg 64

RESULT 38
AAB19865 standard; Protein; 288 AA.
XX AAB19865;
XX AAB19865;
XX 19-MAR-2001 (first entry)
XX Amphotropic MLV surface subunit RBD domain and PHGV motif.
DE Amphotropic; MLV; surface subunit; receptor binding domain; RBD;
KW PHGV motif; activating polypeptide; envelope glycoprotein;
KW retrovirus; fusion defective; infection; cancer; therapy;
KW virucide; cytostatic.
XX Murine leukemia virus.
OS
XX
PN WO200071578-A2.
XX
PD 30-NOV-2000.
XX
PF 19-MAY-2000; 2000WO-EP04534.
XX
XX 20-MAY-1999; 99EP-0401220.
XX
XX (CNRS) CNRS CENT NAT RECH SCI.
PA
PI Cosset F, Lavillette D;
XX
XX WPI; 2001-061346/07.
XX N-PSDB; AAB9065.
XX
XX Use of an activating polypeptide comprising an amino acid sequence from
a surface subunit of a viral envelope glycoprotein, useful for the
rescue of fusion defective viruses or retroviral glycoproteins, and for
treating cancer -
PT
PT Disclosure; Fig 6; 96pp; English.
XX
XX The present sequence is that an activating polypeptide (I),
comprising the receptor binding domain (RBD) and the PHGV
motif of the amphotropic murine leukemia virus surface subunit
(SU), useful in claimed methods of the invention for the rescue of
a fusion defective envelope glycoprotein (II) the glycoprotein of a
sequence corresponding to retrovirus a sequence derived from SU, or
first amino acid of SU encompassing the RBD and the N-terminal part of SU.
XX (II) belongs to a second virus or a second retrovirus of the same
or different type from the first virus or retrovirus. (I) may
recognise the same or different receptors on target cells as are
that is specific to targeted retrovirus. (I) and (II) may be used as a
cells. (I) may be expressed by a host such as a virus or cell, or
(II) may be expressed by a host, such as a virus
incorporating (II) in its envelope, or a cell transformed by for the
nucleotide sequences coding for (II). (I) and (II) are used for
preparation of a drug for the treatment of cancer and for
diseases, and as combined preparation of simultaneous separate
or sequential use for the rescue of the fusion defective property

CC Query Match 43.2%; Score 38; DB 22; Length 288;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLG 12
Db 81 ewdpsdqepvvg 92

RESULT 39
AAP70183
ID AAP70183 standard; Protein; 293 AA.
XX AAP70183;
XX AAP70183;
XX 27-MAY-1991 (first entry)
XX Sequence encoded by signal sequence of S.diastaticus DEX4
DE glucoamylase gene.
KW Yeast expression vector.
KW Yeast expression vector.
XX Saccharomyces diastaticus.
OS
XX EP228254-A.
XX
XX 08-JUL-1987.
XX
XX 18-DEC-1986; 86EP-0309904.
PF
PR 18-DEC-1985; 85US-0810423.
XX
XX (BIOT-) BIOTECHNICA INT INC.
PA
PI Maine GT, Daves RS, Yocum RR;
XX
XX WPI; 1987-186757/27.
DR N-PSDB; AAN70265.
XX
XX Vector for transforming yeast - having secretory signal-encoding
sequenced glucoamylase gene from Saccharomyces diastaticus or
S.cerevisiae
PT
XX Disclosure; Fig 3; 11pp; English.
XX
XX AAN70265 contains the preferred signal sequence of the claimed vector.
The vector can be used to transform yeast cells to produce a desired
protein, eg glucoamylase from A.niger which can be used in brewing,
corn syrup prodn. and grain fermentation for distd. EtOH. prodn.
XX
XX Sequence 293 AA;

Query Match 43.2%; Score 38; DB 8; Length 293;
Best Local Similarity 50.7%; Pred. No. 1.3e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 WDTSAALAPYLGTOEE 16
Db 76 wdssaleewlgrqkk 90

RESULT 40
AAP80770
ID AAP80770 standard; protein; 310 AA.
XX AAP80770;
XX AAP80770;

RESULT	41
AAP90174	
ID	AAP90174 standard; peptide; 314 AA.
XX	
XX	
AC	AAP90174;
XX	
XX	
DT	21-DEC-1989 (first entry)
XX	
DE	Tissue plasminogen activator mutant 2K1.
XX	
XX	
KW	Tissue plasminogen activator; mutant 2K1; fibrinolysis.
XX	
OS	Homo sapiens.
XX	
XX	

```

PN WO8907146-A.
XX
XX
PD 10-AUG-1989.
XX
XX PF 03-FEB-1989; 89WO-US000465.
XX
XX PR 05-FEB-1988; 88US-0152692.
XX
XX PA (INTE-) INTEGRATED GENETICS INC.
XX
XX PI Markland W, Livingston DJ;
XX
XX DR WPI; 1989-249015/34.
XX N-PSDB; AAN90564.
XX
XX PT Rearranged tissue plasminogen activators - prepd. by altering the DNA
XX sequence to introduce AvrII, NheI, SpeI or XbaI cleavage sites.
XX
XX PS Disclosure; Table 10; page 48; 77pp; English.
XX
XX CC Peptide has the kringle 2 domain replaced by another kringle 1 domain
XX It has tPA activity, with extended in vivo half life. Spacing between
XX tPA domains is increased, to increase rate of fibrinolysis or the
XX resistance to inhibition by endogenous tPA inhibitors present in huma
XX plasma. The peptide is used for thrombolysis in the treatment of
XX myocardial infarction, pulmonary embolism, deep vein thrombosis and
XX stroke. See also AAN90564, AAP90168-73, and AAP90175-80.
XX
XX SQ Sequence 314 AA;

Query Match 43.2%; Score 38; DB 10; Length 314;
Best Local Similarity 44.4%; Pred. NO. 1.4e+02;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps

QY 1 DWDTSALA--PYLGTOEE 16
:||||| |||| :
Db 269 nwnsalagkpgysgrlpd 286

RESULT 42
AAW28537
ID ID AAW28537 standard; Protein; 356 AA.
XX
XX AC AAW28537;
XX
XX DT 22-JAN-1998 (first entry)
XX
XX DE mt-PA9.
XX
XX KW Tissue plasminogen activator; chimera; kringle; mt-PA6-D;
KW mt-PA6-E; mt-PA9.
XX
XX OS Homo sapiens.
XX
XX US US5658788-A.
XX
XX PD 19-AUG-1997.
XX
XX XX 22-APR-1991; 91US-0689410.
XX PF 22-APR-1991; 91US-0689410.
XX PR 21-APR-1995; 95US-0427640.
XX
XX PA (ELIL ) LILLY & CO ELI.
XX
XX PI Berg DT, Grinnell BW;
XX
XX DR WPI; 1997-424241/39.
XX N-PSDB; AAT87457.
XX
XX PT Recombinant DNA encoding tissue plasminogen activator variant -
XX comprising signal and pro peptide(s) and chimeric kringle domain,
```

PT produced in di:glycosylated form, useful to treat thromboembolic
 XX diseases
 PS Disclosure: Col 43-46; 44pp; English.
 XX mt-PA6-D (AAW87455) comprises an alternate glycosylation consensus
 CC sequence at the amino acids surrounding position 33 (equivalent
 CC to amino acid 205 of t-PA).
 CC mt-PA6-E (AAW87456) comprises an alternate glycosylation consensus
 CC sequence at the amino acids surrounding position 48 (equivalent
 CC to amino acid 219 of t-PA).
 CC mt-PA9 (AAW87477) is a human t-PA derivative that lacks the finger and
 CC a part of kringle 1 that can be glycosylated, and a part of
 CC kringle 2 that cannot.
 XX Sequence 356 AA;
 SQ

Query Match 43.2%; Score 38; DB 18; Length 356;
 Best Local Similarity 44.4%; Pred. NO. 1.6e+02;
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
 OY 1 DWDTSALA--PYLGTOEE 16
 Db 32 nwnssalackpysgrprd 49

RESULT 43
 ID AAW14432 standard; Protein; 356 AA.
 XX AAW14432;
 AC AAW14432;
 DT 14-MAY-1997 (first entry)
 XX Tissue plasminogen activator derivative, mt-PA9.
 DE Human tissue plasminogen activator; t-PA derivative; mt-PA6; Finger;
 KW Growth factor; kringle 1; kringle 2; monoglycosylated; mt-PA6-E;
 KW mt-PA9; treatment; thromboembolic disorder; consensus sequence;
 KW alternate glycosylation; serine protease.
 XX Synthetic.
 OS
 XX US595736-A.
 PN 21-JAN-1997.
 XX
 XX 22-APR-1991; 91US-0689410.
 XX 22-APR-1991; 91US-0689410.
 PR 27-MAR-1995; 95US-0411260.
 XX (ELIL) LILLY & CO ELI.
 XX Berg DT, Grinnell BW;
 PI WPI; 1997-107532/10.
 DR N-PSDB; AAT62607.
 XX Human tissue plasminogen activator derivs. - useful for treating
 XX thromboembolic disorders
 PS Claim 4; Column 43-46; 35pp; English.
 XX Human tissue plasminogen activator (t-PA) derivs. (especially derivs. of
 CC mt-PA6) have been designed which: (A) lack kringle 1 domain and
 CC kringle 2 domain; (B) lack kringle 2 domain that is mono-
 CC glycosylated; (C) lack amino acid residue equiv. to amino acid 218 of native
 CC human t-PA; (D) lack Finger, Growth Factor and Kringle 1 domains and
 CC comprises a Kringle 2 domain that is monoglycosylated at amino acid 47
 CC (designated mt-PA6-E; AAW14431); and (C) lack Finger and Growth Factor

domains and has chimeric Kringle 1 and 2 domains (designated mt-PA9;
 AAW14432). The t-PA derivatives are useful for treating thromboembolic
 CC disorders. The alternate glycosylation sequence, as well as the
 CC glycosylation consensus sequence within the serine protease domain, is
 CC glycosylated upon the expression and secretion of these molecules from
 CC eukaryotic host cells. Thus, a homogeneous population of di:glycosylated
 CC t-PA derivs. that lack the finger, growth factor and Kringle 1 domains,
 CC is produced. The present sequence is mt-PA9.
 XX Sequence 356 AA;
 SQ

Query Match 43.2%; Score 38; DB 18; Length 356;
 Best Local Similarity 44.4%; Pred. NO. 1.6e+02;
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
 OY 1 DWDTSALA--PYLGTOEE 16
 Db 32 nwnssalackpysgrprd 49

RESULT 44
 ID AAW19868 standard; Protein; 364 AA.
 XX AAW19868;
 AC AAW19868;
 DT 19-MAR-2001 (first entry)
 XX Activating polypeptide A-RBD-epidermal growth factor fusion.
 DE Amphotropic; MLV; surface subunit; receptor binding domain; RBD;
 KW PHOV motif; activating polypeptide; envelope glycoprotein;
 KW retrovirus; fusion defective; infection; cancer; therapy;
 KW virucide; cytosstatic; epidermal growth factor.
 XX Chimeric - Murine leukemia virus.
 OS Chimeric - Mammalia.
 XX WO2000071578-A2.
 PN 30-NOV-2000.
 XX 19-MAY-2000; 2000WO-EP04534.
 XX 20-MAY-1999; 99EP-0401220.
 PR (CNRS) CNRS CENT NAT RECH SCI.
 XX Cosset F, Lavillette D;
 PI WPI; 2001-061346/07.
 DR N-PSDB; AAA89068.
 XX Use of an activating polypeptide comprising an amino acid sequence from
 PT a surface subunit of a viral envelope glycoprotein, useful for the
 PT rescue of fusion defective viruses or retroviral glycoproteins, and for
 XX treating cancer.
 XX Disclosure: Fig 9; 96pp; English.
 PS
 XX The present sequence is that of a fusion protein between an
 CC activating polypeptide comprising the receptor binding domain in
 CC (RBD) and the PHOV motif of a murine leukemia virus
 CC surface subunit and epidermal growth factor (EGF). This is an
 CC surrogate of activating polypeptides (I) used in the invention for the
 CC rescue of a fusion defective envelope glycoprotein (II). (I) may
 CC comprise a sequence corresponding to the SU of an envelope
 CC glycoprotein of a first virus or a first retrovirus, a sequence
 CC derived from SU, or a fragment of SU encompassing the N-terminal
 CC N-terminal part of SU. (II) represents a second virus or a second
 CC retrovirus of the same or different type from the first virus or
 CC retrovirus. (I) may recognise the same or different receptors on

CC target cells as are recognised by (II), and (I) and/or (II) may be
CC linked to a ligand that is specific to target cell receptors, e.g.
CC receptors of tumour cells such as EGF. (I) may be expressed by a
CC host, such as a virus or cell, or by the target cells, transformed
CC by nucleotide sequences coding for (I). (II) may be expressed by a
CC host, such as a virus incorporating (ii) in its envelope, or a cell
CC transformed by nucleotide sequences coding for (II). (I) and (II)
CC are used for the preparation of a drug for the treatment of cancer
CC and infectious diseases, and as combined preparations for
CC simultaneous, separate or sequential use for the rescue of the
CC fusion defective property of (II) or for treatment of cancer
CC pathologies (all claimed).

XX Sequence 364 AA;

Query Match 43.2%; Score 38; DB 22; Length 364;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 DWDTSALAPYLIG 12
Db 157 ewdpsdqepvg 168

RESULT 45

AAR67652
ID AAR67652 standard; Protein; 409 AA.

XX AC AAR67652;

XX DT 16-AUG-1995 (first entry)

XX DE Glucomyrase.

KW Aqualysin I; glucomyrase; extracellular production.

XX OS Saccharomyces cerevisiae.

FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= sig_peptide

XX PN JP06303984-A.

XX PD 01-NOV-1994.

XX PF 19-APR-1993; 93JP-0115378.

XX PR 19-APR-1993; 93JP-0115378.

XX PA (NTSB) JAPAN TOBACCO INC.

XX DR WPI; 1995-018277/03.

XX DR N-PSDB; AAQ75858.

PT Recombinant DNA having sequence deleted from the aqua-lysin I
PT precursor gene - for expression of mature aqua-lysin I in yeast

PS Claim 1; Pages 7-8; 16pp; Japanese.

XX AAQ75858 is the yeast glucomyrase (AAR67652) DNA, into which the
CC aqualysin I (AAR67653) DNA (AAQ75859) minus its C-terminal pro
CC sequence can be inserted. Resulting in the extracellular pro
CC production of aqualysin I in yeast.

SQ Sequence 409 AA;

Query Match 43.2%; Score 38; DB 16; Length 409;
Best Local Similarity 46.7%; Pred. No. 1.9e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 WDTLSALAPYLGTQEE 16
Db 333 wdssaleewlqrqkk 347

Search completed: June 28, 2001, 11:54:15
Job time: 76 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:11 ; Search time 37.02 Seconds
(without alignments)
8.706 Million cell updates/sec

Title: US-09-439-313-562

Perfect score: 88

Sequence: 1 DWDTSALAPYIGTQEE 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 65 summaries

Database :

Issued Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	46	52.3	225	2	US-08-637-759B-127	Sequence 127, App
2	46	52.3	225	3	US-08-871-355A-127	Sequence 127, App
3	41	46.6	233	2	US-08-928-692-31	Sequence 31, Appl
4	39	44.3	270	4	US-09-362-473-10	Sequence 10, Appl
5	39	44.3	429	4	US-09-362-473-12	Sequence 12, Appl
6	38	43.2	43	4	US-08-930-503A-17	Sequence 17, Appl
7	38	43.2	83	2	US-08-811-949-2	Sequence 2, Appl
8	38	43.2	118	4	US-08-983-607-24	Sequence 24, Appl
9	38	43.2	356	1	US-08-427-640-8	Sequence 8, Appl
10	38	43.2	437	2	US-08-811-949-49	Sequence 49, Appl
11	38	43.2	437	2	US-08-811-949-51	Sequence 51, Appl
12	38	43.2	437	2	US-08-811-949-55	Sequence 55, Appl
13	38	43.2	437	2	US-08-811-949-57	Sequence 57, Appl
14	38	43.2	472	2	US-08-811-949-63	Sequence 63, Appl
15	38	43.2	527	1	US-07-609-510B-16	Sequence 16, Appl
16	38	43.2	527	2	US-08-811-949-39	Sequence 39, Appl
17	38	43.2	527	5	PCT-US91-01025A-2	Sequence 2, Appl
18	38	43.2	527	6	5185259-8	Patent No. 5185259
19	38	43.2	527	6	5520913-1	Patent No. 5520913
20	38	43.2	546	6	5200340-6	Patent No. 5200340
21	38	43.2	562	2	US-08-811-949-43	Sequence 43, Appl
22	38	43.2	562	2	US-08-560-098A-50	Sequence 50, Appl
23	38	43.2	562	2	US-08-883-795A-38	Sequence 38, Appl
24	38	43.2	562	6	5185259-3	Patent No. 5185259
25	38	43.2	562	6	5200340-2	Patent No. 5200340
26	38	43.2	562	6	5344773-2	Patent No. 5344773
27	38	43.2	609	3	US-08-941-445A-7	Sequence 7, Appl

RESULT 1
US-08-637-759B-127
; Sequence 127, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

ALIGNMENTS

28	38	43.2	806	1	US-08-270-076A-11	Sequence 11, Appl
29	38	43.2	912	5	PCT-US95-03747-2	Sequence 2, Appl
30	37.5	42.6	260	2	US-08-864-799-3	Sequence 3, Appl
31	37	42.0	121	4	US-08-983-607-20	Sequence 20, Appl
32	37	42.0	200	1	US-08-233-146-1	Sequence 1, Appl
33	37	42.0	200	1	US-08-463-470-1	Sequence 1, Appl
34	37	42.0	442	3	US-08-993-359-26	Sequence 26, Appl
35	36	40.9	588	2	US-08-620-605D-2	Sequence 2, Appl
36	36	40.9	588	1	US-09-005-232A-2	Sequence 2, Appl
37	36	40.9	595	1	US-08-225-989-2	Sequence 2, Appl
38	36	40.9	595	1	US-08-570-923-2	Sequence 2, Appl
39	36	40.9	595	1	US-08-580-014-2	Sequence 2, Appl
40	36	40.9	595	1	US-08-232-087A-2	Sequence 2, Appl
41	36	40.9	595	4	US-09-079-785-2	Sequence 2, Appl
42	35	39.8	149	2	US-08-815-175-3	Sequence 3, Appl
43	35	39.8	205	2	US-08-852-809-5	Sequence 5, Appl
44	35	39.8	316	1	US-08-414-926A-22	Sequence 22, Appl
45	35	39.8	316	2	US-08-926-922-22	Sequence 22, Appl
46	35	39.8	316	3	US-09-253-682-22	Sequence 22, Appl
47	35	39.8	347	3	US-09-059-369-2	Sequence 2, Appl
48	35	39.8	456	2	US-08-860-150-2	Sequence 2, Appl
49	35	39.8	456	3	US-09-338-132-2	Sequence 2, Appl
50	34	38.6	16	1	US-08-446-692-44	Sequence 44, Appl
51	34	38.6	16	2	US-08-488-351A-44	Sequence 44, Appl
52	34	38.6	16	5	PCT-US95-13841-16	Sequence 16, Appl
53	34	38.6	30	1	US-08-446-692-23	Sequence 23, Appl
54	34	38.6	30	2	US-08-488-351A-23	Sequence 23, Appl
55	34	38.6	38	1	US-08-446-692-59	Sequence 59, Appl
56	34	38.6	38	1	US-08-446-692-65	Sequence 65, Appl
57	34	38.6	38	2	US-08-488-351A-59	Sequence 59, Appl
58	34	38.6	38	2	US-08-488-351A-65	Sequence 65, Appl
59	34	38.6	48	1	US-08-446-692-37	Sequence 37, Appl
60	34	38.6	48	2	US-08-488-351A-37	Sequence 37, Appl
61	34	38.6	189	4	US-08-816-346-8	Sequence 8, Appl
62	34	38.6	189	4	US-09-335-411-8	Sequence 8, Appl
63	34	38.6	234	4	US-09-354-129-14	Sequence 14, Appl
64	34	38.6	236	3	US-08-886-269-5	Sequence 5, Appl
65	34	38.6	236	3	US-09-167-647-3	Sequence 3, Appl

NAME: Pabst, Patrea L. 1,284
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET INFORMATION: PMS 101 CON
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 127:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
STANDARDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-127

Query Match 52.3%; Score 46; DB 2; Length 225;
Best Local Similarity 70.0%; Pred. No. 1;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 DWTSAALAPY 10
DB 102 EWDKALAPY 111

RESULT 3
US-08-928-692-31
; Sequence 31, Application US/08928692
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 59587270 No. 5958727/disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 3,728
; REFERENCE/DOCKET INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 233 amino acids
; TYPE: amino acid
; STANDARDNESS: acsingle
; TOPOLOGY: linear
; MOLECULE TYPE: NO. 5958727e
US-08-928-692-31

Query Match 46.5%; Score 41; DB 2; Length 233;
Best Local Similarity 53.8%; Pred. No. 8.4;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 DWTSAALAPYGTQ 14.
DB 35 WDFGALEPYISGQ 47

RESULT 4
US-09-362-473-10
; Sequence 10, Application US/09362473
; Patent No. 6218169
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Richard B.
; APPLICANT: Cahoon, Richard
; APPLICANT: Falcone, S. Carl
; APPLICANT: Morgante, Michele
; APPLICANT: Rafalski, J. Antoni

NAME: Pabst, Patrea L. 1,284
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET INFORMATION: PMS 101
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 127:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acid
STANDARDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-127

Query Match 52.3%; Score 46; DB 2; Length 225;
Best Local Similarity 70.0%; Pred. No. 1;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 DWTSAALAPY 10
DB 102 EWDKALAPY 111

RESULT 2
US-08-871-355A-127
; Sequence 127, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2500 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; FILING DATE: 11-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 127:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acid
; STANDARDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-871-355A-127

; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Aromatic Amino Acid Catabolism Enzymes
; FILE REFERENCE: BB-1197
; CURRENT APPLICATION NUMBER: US/09/362,473
; CURRENT FILING DATE: 1999-07-28
; EARLIER APPLICATION NUMBER: 60/094,783
; EARLIER FILING DATE: JULY 31, 1998
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Zea mays
US-09-362-473-10

Query Match 44.3%; Score 39; DB 4; Length 270;
Best Local Similarity 54.5%; Pred. No. 23;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSALAPYL 12
|:| | |:
Db 97 WETPLGPFLG 107

RESULT 5
US-09-362-473-12

; Sequence 12, Application US/09362473
; Patent No. 6218169
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Edgar B.
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Falco, S. Carl
; APPLICANT: Morgante, Michele
; APPLICANT: Rafalski, J. Antoni
; APPLICANT: Hitz, William D.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Aromatic Amino Acid Catabolism Enzymes
; FILE REFERENCE: BB-1197
; CURRENT APPLICATION NUMBER: US/09/362,473
; CURRENT FILING DATE: 1999-07-28
; EARLIER APPLICATION NUMBER: 60/094,783
; EARLIER FILING DATE: JULY 31, 1998
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-362-473-12

Query Match 44.3%; Score 39; DB 4; Length 429;
Best Local Similarity 54.5%; Pred. No. 39;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSALAPYL 12
|:| | |:
Db 255 WETPLGPFLG 265

RESULT 6
US-08-930-503A-17

; Sequence 17, Application US/08930503A
; Patent No. 6132731
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MURINE LEUKEMIA VIRUS VECTORS (AS AMENDED)
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
; STREET: 2033 K Street, N.W. - Suite 800

; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,503A
; FILING DATE: October 8, 1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9506782.3
; FILING DATE: 01-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/GB96/00776
; FILING DATE: 01-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:

; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-930-503A-17

Query Match 43.2%; Score 38; DB 4; Length 43;
Best Local Similarity 50.0%; Pred. No. 4;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 12
|:| | |:
Db 4 EWDPDQEPVG 15

RESULT 7
US-08-811-949-2

; Sequence 2, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; Sequence 49, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-49

Query Match 43.2%; Score 38; DB 2; Length 437;
Best Local Similarity 44.4%; Pred. No. 60;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

Qy 1 DWDTSALA--PYLGTOEE 16
Db 25 MNSSALAKPYSGRRPD 42

RESULT 11
US-08-811-949-51
; Sequence 51, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA

; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-51

Query Match 43.2%; Score 38; DB 2; Length 437;
Best Local Similarity 44.4%; Pred. No. 60;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

Qy 1 DWDTSALA--PYLGTOEE 16
Db 25 MNSSALAKPYSGRRPD 42

RESULT 12
US-08-811-949-55
; Sequence 55, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: NOTANI, JOUJI
; APPLICANT: KOBAYASHI, MASAKAZU
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-51

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-55

Query Match 43.2%; Score 38; DB 2; Length 437;
Best Local Similarity 44.4%; Pred. No. 60;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
Db 25 NWNSSALAKPYSGRRPD 42

RESULT 13
US-08-811-949-57
; Sequence 57, Application US/08811949
; Patent No. 5840533
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAKAI, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: KOBAYASHI, MASAKAZU
; APPLICANT: KOBAYASHI, JOUJI
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P.C. JEFFERSON DAVIS HIGHWAY, SUITE 400
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/811.949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-811-949-57

Query Match 43.2%; Score 38; DB 2; Length 437;
Best Local Similarity 44.4%; Pred. No. 60;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
Db 25 NWNSSALAKPYSGRRPD 42

RESULT 14
US-08-811-949-59
; Sequence 59, Application US/07609510B
; Patent No. 542700
; GENERAL INFORMATION:
; APPLICANT: Berg et al.
; TITLE OF INVENTION: Method for Altering Post-Translational Processing of Tissue
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 Mb storage
; SOFTWARE: Macintosh
; OPERATING SYSTEM: Macintosh
```

Query Match 43.2%; Score 38; DB 2; Length 527;
Best Local Similarity 44.4%; Pred. No. 74;
Matches 8; Conservative 5; Mismatches 3; Indels

```

* NUMBER OF SEQUENCES: 15
* CURRENT APPLICATION DATA:
* APPLICATION NUMBER: US/07/489,855
* FILING DATE: 02-MAR-1990
* PRIOR APPLICATION DATA:
* APPLICATION NUMBER: 12,694
* FILING DATE: 09-FEB-1987
* APPLICATION NUMBER: 483,052

```


; FILING DATE: 07-APR-1983
; APPLICATION NUMBER: 398,003
; FILING DATE: 14-JUL-1982
; APPLICATION NUMBER: 374,860
; FILING DATE: 05-MAY-1982
; SEQ ID NO:6;
; LENGTH: 527
5185259-8

Query Match
Best Local Similarity 43.2%; Score 38; DB 6; Length 527;
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
DB 115 WNSSSALAKPYSGRRPD 132
:::|||||

RESULT 19

5520913-1
; Patent No. 5520913
; APPLICANT: ANDERSON, STEPHEN; BENNETT, WILLIAM F.; BOTSTEIN,
; DAVID; HIGGINS, DEBORAH L.; PAONI, NICHOLAS F.; ZOLLER, MARK J.
; SYMOGIC PROPERTIES
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR HAVING
; CURRENT APPLICANT DATA:
; FILING DATE: 06-JUL-1993
; PRIOR APPLICATION NUMBER: US/88,451
; APPLICATION NUMBER: 770,510
; FILING DATE: 03-OCT-1991
; APPLICATION NUMBER: 240,856
; FILING DATE: 24-JUL-1989
; APPLICATION NUMBER: 240,856
; FILING DATE: 02-SEP-1988
; SEQ ID NO:1;
; LENGTH: 527
5520913-1

Query Match
Best Local Similarity 43.2%; Score 38; DB 6; Length 527;
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
DB 115 WNSSSALAKPYSGRRPD 132
:::|||||

RESULT 20

5200340-6
; Patent No. 5200340
; APPLICANT: ROESTER, DONALD C.; MULVIHILL, EILEEN R.; O'HARA,
; PATRICK J.; PINGEL, KURT; YOSHITAKE, SHINJI
; TITLE OF INVENTION: THROMBIN-ACTIVATED TISSUE PLASMINOGEN
; ACTIVATORS
; NUMBER OF SEQUENCES: 34
; CURRENT APPLICATION DATA:
; FILING DATE: 22-MAY-1987
; SEQ ID NO:6;
; LENGTH: 546
5200340-6

Query Match
Best Local Similarity 43.2%; Score 38; DB 6; Length 546;
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
:::|||||

Db 150 WNSSSALAKPYSGRRPD 167

RESULT 21

US-08-811-949-43
; Sequence 43560098A
; Patent No. 43560098A
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINEO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: SASAKI, HITOSHI
; APPLICANT: HAYASHI, MASAKO
; APPLICANT: KOTANI, KENJI
; APPLICANT: KOSHI, SHINJI
; APPLICANT: KOSHI, SHINJI
; TITLE OF INVENTION: TISSUE PLASMINOGEN ACTIVATOR
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ATTORNEYS: P.C. JEFFERSON DAVIS HIGHWAY, SUITE 400
; STREET: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; CURRENT APPLICANT DATA:
; FILING DATE: 05-MAR-1997
; APPLICATION NUMBER: US/08/811,949
; FILING DATE: 05-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-966-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-811-949-43

Query Match
Best Local Similarity 43.2%; Score 38; DB 2; Length 562;
Matches 8; Conservative 5; Mismatches 3; Indels 3; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
DB 150 WNSSSALAKPYSGRRPD 167
:::|||||

RESULT 22

US-08-560-098A-50
; Sequence 50560098A
; Patent No. 50560098A
; GENERAL INFORMATION:
; APPLICANT: WNEEDT, Stephan
; APPLICANT: HEINZEL-WIELAND, Regina
; APPLICANT: STEFFENS, Gerd Josef
; TITLE OF INVENTION: Proteins having Fibrinolytic and
; TITLE OF INVENTION: Coagulation-inhibiting Properties
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: EDWARDS, MCKEOWN, Edwards & Lenahan
; STREET: 1200 G Street, N.W., Suite 700
; CITY: Washington

```
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/560,098A
; FILING DATE: 17-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P 44 40 892.7
; FILING DATE: 17-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: EVANS, Joseph D.
; REGISTRATION NUMBER: 26,269
; REFERENCE/DOCKET NUMBER: 148/42448
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-8800
; TELEFAX: (202) 628-8844
; INFORMATION FOR SEQ ID NO: 50:
; LENGTH: 562 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-560-098A-50

Query Match 43.2%; Score 38; DB 2; Length 562;
Best Local Similarity 44.4%; Pred. No. 80;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
Db 150 WNSSALAQKPYSGRRPD 167

RESULT 23
US-08-883-795A-38
; Sequence 38, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcuve, Genevieve
; APPLICANT: Awang, Gregor
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
```

```
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein (tpa)
; US-08-883-795A-38

Query Match 43.2%; Score 38; DB 2; Length 562;
Best Local Similarity 44.4%; Pred. No. 80;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
Db 150 WNSSALAQKPYSGRRPD 167

RESULT 24
5185259-3
; Patent No. 5185259
; APPLICANT: GOEDDEL, DAVID V.; KOHR, WILLIAM J.; PENNICA, DIANE;
; VEHAR, GORDON A.
; TITLE OF INVENTION: TRUNCATED HUMAN TISSUE PLASMINOGEN
; ACTIVATOR
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/489,855
; FILING DATE: 02-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 12,694
; FILING DATE: 09-FEB-1987
; APPLICATION NUMBER: 483,052
; FILING DATE: 07-APR-1983
; APPLICATION NUMBER: 398,003
; FILING DATE: 14-JUL-1982
; APPLICATION NUMBER: 374,860
; FILING DATE: 05-MAY-1982
; SEQ ID NO: 3
; LENGTH: 562
; 5185259-3

Query Match 43.2%; Score 38; DB 6; Length 562;
Best Local Similarity 44.4%; Pred. No. 80;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;

QY 1 DWDTSALA--PYLGTQEE 16
Db 150 WNSSALAQKPYSGRRPD 167

RESULT 25
5200340-2
; Patent No. 5200340
; APPLICANT: FOSTER, DONALD C.; MULVIHILL, EILEEN R.; O'HARA,
; PATRICK J.; PINGEL, KURT; YOSHITAKE, SHINJI
; TITLE OF INVENTION: THROMBIN-ACTIVATED TISSUE PLASMINOGEN
; ACTIVATORS
; NUMBER OF SEQUENCES: 34
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/53,412
; FILING DATE: 22-MAY-1987
; SEQ ID NO: 2
; LENGTH: 562
; 5200340-2

Query Match 43.2%; Score 38; DB 6; Length 562;
Best Local Similarity 44.4%; Pred. No. 80;
Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
```

```
QY 1 DWTSAALA--PYLGTOEE 16
; :|:|:|:|:|
; MOLECULE TYPE: Protein
; US-08-941-445A-7
;
; TOPOLOGY: linear
;
; Query Match
; Best Local Similarity 43.2%; Score 38; DB 3; Length 609;
; Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 DWTSAALAPYL 11
; | | | | |
; Db 234 DWTGTPALSYL 244
;
; RESULT 28
; US-08-270-076A-11
; Sequence 11, Application US/08270076A
; Patent No. 5667986
; GENERAL INFORMATION:
; APPLICANT: Sleep, Darrell
; APPLICANT: Goodey, Andrew R
; INVENTOR: Sleep, Darrell
; TITLE OF INVENTION: Yeast Promoter
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The BOC Group, Inc.
; STREET: 100 Mountain Avenue, Murray Hill
; STATE: New Providence
; COUNTRY: USA
; ZIP: 07974
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/270.076A
; FILING DATE: 01-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8923521.2
; FILING DATE: 18-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/597,687
; FILING DATE: 16-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/925,286
; FILING DATE: 01-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Swope, R. Hain
; REGISTRATION NUMBER: 24864
; REFERENCE/DOCKET NUMBER: 92H834-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908/771-6292
; TELEFAX: 908/771-6293
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 806 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-270-076A-11
;
; Query Match
; Best Local Similarity 43.2%; Score 38; DB 1; Length 806;
; Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
;
; QY 2 DWTSAALAPYLTOEE 16
; | | | | |
; Db 333 WDSSALEWLRQKK 347
;
; Query Match
; Best Local Similarity 46.7%; Score 38; DB 1; Length 806;
; Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
;
; QY 1 DWTSAALA--PYLGTOEE 16
; :|:|:|:|:|
; MOLECULE TYPE: Protein
; US-08-941-445A-7
;
; TOPOLOGY: linear
;
; Query Match
; Best Local Similarity 43.2%; Score 38; DB 6; Length 562;
; Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
;
; QY 1 DWTSAALA--PYLGTOEE 16
; | | | | |
; Db 150 WNSSSALAQRPGRRPD 167
;
; RESULT 27
; US-08-941-445A-7
; Sequence 7, Application US/08941445A
; Patent No. 6107060
; GENERAL INFORMATION:
; APPLICANT: Keeling, Peter
; APPLICANT: Guan, Hanping
; TITLE OF INVENTION: Starch Encapsulation
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle
; CITY: Boulder
; STATE: CO
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/941.445A
; FILING DATE: 30-SEP-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,855
; FILING DATE: 30-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Winner and Sullivan P 28,547
; REGISTRATION NUMBER: 89-97
; REFERENCE/DOCKET NUMBER: 89-97
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 606 amino acids
; TYPE: amino acid
```

RESULT 29
PCT-US95-03747-2
; Sequence 2, Application PC/TUS9503747
; GENERAL INFORMATION:
; APPLICANT: LA JOLLA CANCER RESEARCH FOUNDATION
; TITLE OF INVENTION: Brevican, A Glial Cell Proteoglycan
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/03747
; FILING DATE: 27-MAR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Imbra, Richard J.
; REGISTRATION NUMBER: 37,643
; REFERENCE/DOCKET NUMBER: FP-LJ 1453
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 912 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-03747-2

Query Match 43.2%; Score 38; DB 5; Length 912;
Best Local Similarity 37.5%; Pred. No. 1.4e+02;
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQEE 16
Db 433 EFQTOSIVPLGSSEE 448
RESULT 30
US-08-864-799-3
; Sequence 3, Application US/08864799
; Patent No. 5858714
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN METAXIN PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,799

; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0312 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 289739
US-08-864-799-3

Query Match 42.6%; Score 37.5; DB 2; Length 260;
Best Local Similarity 53.3%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1 DW-DTSALAPYLGTQ 14
Db 22 DWEDVSLFTPLNDQ 36

RESULT 31
US-08-983-607-20
; Sequence 20, Application US/08983607
; Patent No. 6140470
; GENERAL INFORMATION:
; APPLICANT: Alan Garen
; APPLICANT: Xiaohong Cai
; TITLE OF INVENTION: Human Anti-Tumor Monoclonal Anti-
; TITLE OF INVENTION: bodies
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Department of Molecular Biophysics
; ADDRESSEE: and Biochemistry, Yale University
; STREET: 266 Whitney Avenue
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: United States of America
; ZIP: 06520-8114
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" 1.44 Mb diskette
; COMPUTER: IBM PC
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Word Processing
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/983,607
; FILING DATE: April 27, 1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/IB96/01032
; FILING DATE: June 28, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary M. Krinsky
; REGISTRATION NUMBER: 32423
; REFERENCE/DOCKET NUMBER: OCR-679
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203-773-9544
; TELEFAX: 203-773-1183
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 121 residues
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: linear
; ORIGINAL SOURCE: polypeptide
; DESCRIPTION: Homo sapiens (melanoma patient immu-
; ORGANISM: Homo sapiens (melanoma patient tumor cells)
; INDIVIDUAL ISOLATE: peripheral blood lymphocytes
; IMMEDIATE SOURCE: DM14 scfv antibodies obtained from
; LIBRARY: FUSE5 fusion phage construct
; CLONE: H18
; FEATURE KEY: heavy chain
;
US-08-983-607-20

Query Match 42.0%; Score 37; DB 4; Length 121;
Best Local Similarity 60.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSAAPYL 11
DB 72 WDTSLTAYM 81

RESULT 32
US-08-233-146-1
; Sequence 1, Application US/08233146
; Patent No. 5648256
; GENERAL INFORMATION:
; APPLICANT: BEPPU, TERUHIKO
; APPLICANT: YAMADA, HIDEAKI
; APPLICANT: NAGASAWA, TORU
; APPLICANT: HORINOUCHI, SUHARU
; APPLICANT: NISHIYAMA, MAKOTO
; TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
; TITLE OF INVENTION: NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
; TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/233,146
; APPLICATION NUMBER: 435
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,746
; FILING DATE: 02-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 18,872
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 121 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: linear
; ORIGIN: SOURCE: pseudomonas chlororaphis
; STRAIN: B23 (FERM BP-187)
;
US-08-983-607-20

; LENGTH: 121 residues
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: linear
; ORIGINAL SOURCE: polypeptide
; DESCRIPTION: Homo sapiens (melanoma patient immu-
; ORGANISM: Homo sapiens (melanoma patient tumor cells)
; INDIVIDUAL ISOLATE: peripheral blood lymphocytes
; IMMEDIATE SOURCE: DM14 scfv antibodies obtained from
; LIBRARY: FUSE5 fusion phage construct
; CLONE: H18
; FEATURE KEY: heavy chain
;
US-08-983-607-20

Query Match 42.0%; Score 37; DB 1; Length 200;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WDTSAAPYL 11
DB 156 WDTSAESRYL 165

RESULT 33
US-08-463-470-1
; Sequence 1, Application US/08463470
; Patent No. 5789211
; GENERAL INFORMATION:
; APPLICANT: BEPPU, TERUHIKO
; APPLICANT: YAMADA, HIDEAKI
; APPLICANT: NAGASAWA, TORU
; APPLICANT: HORINOUCHI, SUHARU
; APPLICANT: NISHIYAMA, MAKOTO
; TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
; TITLE OF INVENTION: NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
; TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/463,470
; APPLICATION NUMBER: 435
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,146
; FILING DATE: 22-APR-1994
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 18,872
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: linear
; ORIGIN: SOURCE: pseudomonas chlororaphis
; STRAIN: B23 (FERM BP-187)
;
US-08-463-470-1
```

Query Match 42.0%; Score 37; DB 1; Length 200;
Best Local Similarity 70.0%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WDTLSALAPYL 11
| | | | : | |
DB 156 WDTSAESRYL 165

RESULT 34
US-08-993-359-26
; Sequence 26, Application US/08993359A
; Patent No. 6039942
; GENERAL INFORMATION:
; APPLICANT: Lassen, Soren F.
; APPLICANT: Bech, Lisbeth
; APPLICANT: Ohmann, Anders
; APPLICANT: Breinholt, Jens
; APPLICANT: Fuglsang, Claus C.
; APPLICANT: Ostergaard, Peter R.
; TITLE OF INVENTION: Phytase Polypeptides
; FILE REFERENCE: 5383.500-US
; CURRENT APPLICATION NUMBER: US/08/993,359A
; CURRENT FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 1480/96
; EARLIER FILING DATE: 1996-12-20
; EARLIER APPLICATION NUMBER: 1481/96
; EARLIER FILING DATE: 1996-12-20
; EARLIER APPLICATION NUMBER: 0301/97
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: 0529/97
; EARLIER FILING DATE: 1997-05-07
; EARLIER APPLICATION NUMBER: 1388/97
; EARLIER FILING DATE: 1997-12-01
; EARLIER APPLICATION NUMBER: 60/046,082
; EARLIER FILING DATE: 1997-05-09
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26
; LENGTH: 442
; TYPE: PRT
; ORGANISM: *Paxillus involutus*
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(19)
US-08-993-359-26

Query Match 42.0%; Score 37; DB 3; Length 442;
Best Local Similarity 54.5%; Pred. No. 91;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTLSALAPYL 12
| | | | : | |
DB 367 WDTSSLVPFSG 377

RESULT 35
US-08-620-605D-2
; Sequence 2, Application US/08620605D
; Patent No. 5846811
; GENERAL INFORMATION:
; APPLICANT: SHIBATANI, TAKEJI
; APPLICANT: AKATSUKA, HIROYUKI
; APPLICANT: KAWAI, ERI
; TITLE OF INVENTION: GENE PARTICIPATING IN THE MECHANISM OF
; TITLE OF INVENTION: SECRETION OF ESTERASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH

STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/620,605D
FILING DATE: 22-MAR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 0020-3955
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)-205-8000
TELEFAX: (703)-205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 588 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-620-605D-2

Query Match 40.9%; Score 36; DB 2; Length 588;
Best Local Similarity 45.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTLSALAPYL 12
| | | | : | |
DB 399 WNKDELGPYIG 409

RESULT 36
US-09-005-232A-2
; Sequence 2, Application US/09005232A
; Patent No. 5922568
; GENERAL INFORMATION:
; APPLICANT: SHIBATANI, TAKEJI
; APPLICANT: AKATSUKA, HIROYUKI
; APPLICANT: KAWAI, ERI
; TITLE OF INVENTION: GENE PARTICIPATING IN THE MECHANISM OF
; TITLE OF INVENTION: SECRETION OF ESTERASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,232A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)-205-8000
TELEFAX: (703)-205-8050
INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 588
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-005-232A-2

Query Match      40.9%; Score 36; DB 2; Length 588;
Best Local Similarity 45.5%; Pred. NO. 1.9e+02;
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 WDTSLAPYLIG 12
DB 399 WNKDELGLPIIG 409

RESULT 37
US-08-225-989-2
; Sequence 2, Application US/08225989
; Patent No. 5480910
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Smith, Craig A.
; APPLICANT: Gruss, Hans-Jurgen
; TITLE OF INVENTION: No. 5480910el Cytokine That Binds CD30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,989
; FILING DATE: 12 APRIL 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/966,775
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 907,224
; FILING DATE: 01-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,660
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 892,459
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE CITATION NUMBER: 2804-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

; SEQUENCE CHARACTERISTICS:
; LENGTH: 588
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-225-989-2

Query Match      40.9%; Score 36; DB 1; Length 595;
Best Local Similarity 57.1%; Pred. NO. 1.9e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 DTSALAPYLGTQEE 16
DB 321 DTTFEAPPLGTQPD 334

RESULT 38
US-08-570-923-2
; Sequence 2, Application US/08570923
; Patent No. 5677430
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Smith, Craig A.
; APPLICANT: Gruss, Hans-Jurgen
; TITLE OF INVENTION: No. 5677430el Cytokine That Binds CD30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/570,923
; FILING DATE: 12-DEC-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,989
; FILING DATE: 12 APRIL 1994
; APPLICATION NUMBER: US 07/966,775
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 907,224
; FILING DATE: 01-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,660
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 892,459
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE CITATION NUMBER: 2804-E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

US-08-570-923-2

Query Match 40.9%; Score 36; DB 1; Length 595;
Best Local Similarity 57.1%; Pred. No. 1.9e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 16
||: || ||||| :
DB 321 DTTFEAPPLGTQPD 334

RESULT 39

US-08-580-014-2
; Sequence 2, Application US/08580014
; Patent No. 5753203
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Armitage, Richard J.
; APPLICANT: Gruss, Hans-Jurgen
; TITLE OF INVENTION: No. 5753203el Cytokine That Binds CD30
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathryn A. Seese, Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple 7.1
; SOFTWARE: Microsoft Word, Version 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/580,014
; FILING DATE: 20-DEC-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/225,989
; FILING DATE: 12 APRIL 1994
; APPLICATION NUMBER: US 07/966,775
; FILING DATE: 27-OCT-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 907,224
; FILING DATE: 01-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,660
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 892,459
; FILING DATE: 02-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 889,717
; FILING DATE: 26-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Seese, Kathryn A.
; REGISTRATION NUMBER: 32,172
; REFERENCE/DOCKET NUMBER: 2804-E
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; TELEX: 756822
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-580-014-2

Query Match 40.9%; Score 36; DB 1; Length 595;
Best Local Similarity 57.1%; Pred. No. 1.9e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 16
||: || ||||| :
DB 321 DTTFEAPPLGTQPD 334

RESULT 40

US-08-232-087A-2
; Sequence 2, Application US/08232087A
; Patent No. 5866372
; GENERAL INFORMATION:
; APPLICANT: Stein, Harald
; APPLICANT: D rkop, Horst
; APPLICANT: Latza, Ute
; TITLE OF INVENTION: Lymphoid CD30-Antigen
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,087A
; FILING DATE: 08-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 756-103P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-232-087A-2

Query Match 40.9%; Score 36; DB 2; Length 595;
Best Local Similarity 57.1%; Pred. No. 1.9e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 16
||: || ||||| :
DB 321 DTTFEAPPLGTQPD 334

RESULT 41
US-09-079-785-2
; Sequence 2, Application US/09079785
; Patent No. 6143869
; GENERAL INFORMATION:
; APPLICANT: Goodwin, Raymond G.
; APPLICANT: Smith, Craig A.
; APPLICANT: Armitage, Richard J.
; APPLICANT: Gruss, Hans-Jurgen
; TITLE OF INVENTION: No. 6143869el Cytokine That Binds CD30
; NUMBER OF SEQUENCES: 23


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: A. Seese, Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM: disk
MEDIUM TYPE: IBM Compatible
OPERATING SYSTEM: Apple Macintosh
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,785
CLASSIFICATION:
PRIORITY:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/225,989
FILING DATE: 12 APRIL 1994
APPLICATION NUMBER: US 07/966,775
FILING DATE: 27-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 907,224
FILING DATE: 01-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,660
FILING DATE: 15-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 892,459
FILING DATE: 02-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 889,717
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hillman, Jennifer L.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2804-E
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 595 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-079-785-2

```

```

Query Match 40.9% Score 36; DB 4; Length 595;
Best Local Similarity 57.1% Pred. No. 1.9e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0;

```

```

Qy 3 DTSALAPVLTQEE 16
Db 321 DTTPEAPPLQTQPD 334

```

```

RESULT 42
US-08-815-175-3
; Sequence 3, Application US/08815175
; Patent No. 5831052
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL PROLINE-RICH ACIDIC PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Pharmaceuticals, Inc.
; STREET: 3174 Potter Drive

```

```

CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FAST-SEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/815,175
FILING DATE: Filed Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0225 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-1025
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 149 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: near
IMMUNOGENICITY:
LIBRARY: GenBank
CLONE: 899433
US-08-815-175-3

```

```

Query Match 39.8% Score 35; DB 2; Length 149;
Best Local Similarity 75.0% Pred. No. 58;
Matches 6; Conservative 0; Mismatches 2; Indels 0;

```

```

Qy 2 WDTSLALAP 9
Db 45 WDTRALEP 52

```

```

RESULT 43
US-08-852-809-5
; Sequence 5, Application US/08852809
; Patent No. 5831052
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NEW HUMAN TRANSLOCATION ASSOCIATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cytoc Pharmaceuticals, Inc.
; STREET: 3174 Potter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FAST-SEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/852,809
FILING DATE: Herewith
CLASSIFICATION: 514
TELECOMMUNICATION DATA:
APPLICATION NUMBER:

```

```
;
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0238 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1063411
; US-08-852-809-5

Query Match 39.8%; Score 35; DB 2; Length 205;
Best Local Similarity 46.2%; Pred. No. 84;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 TSALAPYLGTQEE 16
| : : | | | : | |
Db 61 TDSIGPYLSKEE 73

RESULT 44
US-08-414-926A-22
; Sequence 22, Application US/08414926A
; Patent No. 5721354
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-2155
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,926A
; FILING DATE: March 31, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.16
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..316
; OTHER INFORMATION: /label= UL148
; US-08-926-922-22

Query Match 39.8%; Score 35; DB 2; Length 316;
Best Local Similarity 37.5%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTOEE 16
| | | : | : |
Db 60 DWDSMHCTPFWSTDLE 75

RESULT 45
US-08-926-922-22
; Sequence 22, Application US/08926922
; Patent No. 5925751
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cseri Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,922
; FILING DATE: September 10, 1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.16
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..316
; OTHER INFORMATION: /label= UL148
; US-08-926-922-22

Query Match 39.8%; Score 35; DB 2; Length 316;
Best Local Similarity 37.5%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTOEE 16
| | | : | : |
Db 60 DWDSMHCTPFWSTDLE 75

Search completed: June 28, 2001, 11:56:12
Job time: 193 sec
```

```
;
; OTHER INFORMATION: /label= UL148
; US-08-414-926A-22

Query Match 39.8%; Score 35; DB 1; Length 316;
Best Local Similarity 37.5%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTOEE 16
| | | : | : |
Db 60 DWDSMHCTPFWSTDLE 75

RESULT 45
US-08-926-922-22
; Sequence 22, Application US/08926922
; Patent No. 5925751
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cseri Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,922
; FILING DATE: September 10, 1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 316 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.16
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..316
; OTHER INFORMATION: /label= UL148
; US-08-926-922-22

Query Match 39.8%; Score 35; DB 2; Length 316;
Best Local Similarity 37.5%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTOEE 16
| | | : | : |
Db 60 DWDSMHCTPFWSTDLE 75

Search completed: June 28, 2001, 11:56:12
Job time: 193 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:59 ; Search time 42.81 Seconds
(without alignments)
28.470 Million cell updates/sec

Title: US-09-439-313-562

Perfect score: 88

Sequence: 1 DWDTSALAPYLGTQEE 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database :

PIR_68:**

1: pir1:**

2: pir2:**

3: pir3:**

4: pir4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	51.1	685	G82022	methionine--trna 1
2	45	51.1	685	A81246	methionyl--trna syn
3	45	51.1	831	T00323	chitinase (EC 3.2.
4	45	51.1	847	F83517	probable penicilli
5	43	48.9	275	D83536	hypothetical prote
6	43	48.9	402	S74332	hypothetical prote
7	43	48.9	459	T35317	probable serine/th
8	43	48.9	626	E70747	hypothetical prote
9	42	47.7	207	S15205	superoxide dismuta
10	42	47.7	490	C81748	DNA gyrase, chain
11	42	47.7	643	S00480	potassium channel
12	42	47.7	656	JH0193	potassium channel
13	42	47.7	4485	T08044	dynamin gamma heavy
14	41	46.6	217	E71098	probable aroM prot
15	41	46.6	233	DSBYN	superoxide dismuta
16	41	46.6	343	T31269	4-hydroxy-2-oxoval
17	41	46.6	415	C69882	gamma-glutamyl pho
18	41	46.6	427	D72528	probable L-asparag
19	40	45.5	207	S06599	superoxide dismuta
20	40	45.5	258	T31723	hypothetical prote
21	40	45.5	294	F83371	hypothetical prote
22	40	45.5	484	D65230	hypothetical prote
23	40	45.5	484	A86116	hypothetical prote
24	40	45.5	492	T38093	probable protein d
25	40	45.5	495	D86442	probable amino aci
26	40	45.5	732	C84487	hypothetical prote
27	40	45.5	957	E84547	probable disease r
28	40	45.5	1265	T51314	probable CO-induce
29	39.5	44.9	575	E84184	hypothetical prote

30	39	44.3	81	2	A70825	hypothetical prote
31	39	44.3	175	2	G70823	hypothetical prote
32	39	44.3	320	2	S73413	transcription anti
33	39	44.3	337	2	T32099	hypothetical prote
34	39	44.3	346	2	C71103	probable aspartoki
35	39	44.3	347	2	D84333	alcohol dehydrogen
36	39	44.3	387	2	S75981	magnesium/cobalt t
37	39	44.3	453	2	A65031	biphenyl dioxygena
38	39	44.3	453	2	G85898	hypothetical prote
39	39	44.3	453	2	A69494	glu-tRNA amidotran
40	39	44.3	504	2	H69333	medium-chain acyl-
41	39	44.3	580	2	G83339	hypothetical prote
42	39	44.3	677	1	SYECMT	methionine--trna 1
43	39	44.3	677	2	C85839	methionine trna syn
44	39	44.3	677	2	C83210	methionyl--trna 1
45	39	44.3	682	2	I64113	methionine--trna 1
46	39	44.3	702	2	C82792	methionyl--trna syn
47	39	44.3	727	2	A56879	diacylglycerol kin
48	39	44.3	731	2	D82249	methionyl--trna syn
49	39	44.3	734	1	S09156	diacylglycerol kin
50	39	44.3	735	2	S12969	diacylglycerol kin
51	39	44.3	948	2	G83264	hypothetical prote
52	38	43.2	224	2	T04447	hypothetical prote
53	38	43.2	247	2	S74902	aquaporin z - Syne
54	38	43.2	256	2	T01189	anthocyanin biosyn
55	38	43.2	264	2	T03974	anthocyanin biosyn
56	38	43.2	266	2	T03972	anthocyanin biosyn
57	38	43.2	267	2	T01188	anthocyanin biosyn
58	38	43.2	268	2	T03729	anthocyanin biosyn
59	38	43.2	271	2	T03715	anthocyanin biosyn
60	38	43.2	277	2	F85530	carboxylesterase (
61	38	43.2	321	2	T38427	probable glucan 1,
62	38	43.2	541	2	T34701	hypothetical prote
63	38	43.2	544	2	B75379	CTP synthase - Del
64	38	43.2	562	1	UKHUT	t-plasminogen acti
65	38	43.2	609	2	JQ0703	UDPglucose--starch

ALIGNMENTS

RESULT 1

G82022
methionine--trna ligase (EC 6.1.1.10) NMA0275 [imported] - Neisseria meningitidis (st
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: G82022
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: A81775; MUID:20222556
A:Accession: G82022
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-685 <PAR>
A:Cross-references: GB:A162752; GB:AL157959; NID:g7378778; PIDN:CAB83583.1; PID:g737
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: metG; NMA0275
C:Superfamily: methionine--trna ligase
C:Keywords: ligase

Query Match 51.1%; Score 45; DB 2; Length 685;

Best Local Similarity 66.7%; Pred. No. 13;

Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 DWDTSALAPYL 12

Db 237 DWDTSALAPYFG 248

Nature 406, 959-964, 2000

A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen.

A:Reference number: AB2950; MUID:20437337

A:Accession: F83517

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-847 <STO>

A:Cross-references: GB:AE004535; GB:AE004091; NID:g9946936; PIDN:AAG04421.1; GSPDB:GN:G000000000

C:Experimental source: Strain PAO1

C:Genetics:

A:Gene: PA1032

Query Match 51.1%; Score 45; DB 2; Length 847;
Best Local Similarity 52.9%; Pred. NO. 16;
Matches 9; Conservative 3; Mismatches 2; Gaps 1;

OY 1 DW-DTSALAPYLGTQEE 15
 : | : | : | : | : |
DB 721 EWQSDSKMAPYLGAGE 737

RESULT 5

D83536
C:Species: Pseudomonas aeruginosa (strain PAO1)
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
R:Accession: D83536
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; L.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, K.R.; Kas, A.; Larbig, K.; L.
Lory, S.; Olson, S.G.; Olson, M.V.
A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen.

A:Reference number: AB2950; MUID:20437337

A:Accession: D83536

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-275 <STO>

A:Cross-references: GB:AE004532; GB:AE004091; NID:g9946768; PIDN:AAG04267.1; GSPDB:GN:G000000000

C:Experimental source: Strain PAO1

C:Genetics:

A:Gene: PA0878

Query Match 48.9%; Score 43; DB 2; Length 275;
Best Local Similarity 56.7%; Pred. NO. 10;
Matches 8; Conservative 2; Mismatches 0; Gaps 0;

OY 5 SALAPYLGTQEE 16
 | | | | : | | |
DB 4 SAFAPWIGROEE 15

RESULT 6

S24332
C:Species: *Synechocystis* sp. (strain PCC 6803)
C:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, J.;
DNA "S. Okumura, S. Wada, I.; Watanabe, A.; Yamada, K.; Tas
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.

A:Reference number: S74322; MUID:97061201

A:Accession: S74332

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-408 <NN>

A:Cross-references: EMBL:D64001; GB:AB001339; NID:gl001102; PIDN:BAAL0250.1; PID:d101

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 48.9%; Score 43; DB 2; Length 402;
 Best Local Similarity 53.8%; Pred. No. 16;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQ 14
 I: I:||||| I
 DB 6 WENFAVPYAGEQ 18

RESULT 7
 T35317
 probable serine/threonine protein kinase - Streptomyces coelicolor (fragment)
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T35317
 R:Oliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, May 1999
 A:Reference number: 221575
 A:Accession: T35317
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-459 <COL>
 A:Cross-references: EMBL:AL049863; PIDN:CAB42926.1; GSPDB:GN00070; SCOEDB:SC5H1.01
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC5H1.01

Query Match 48.9%; Score 43; DB 2; Length 459;
 Best Local Similarity 58.8%; Pred. No. 18;
 Matches 10; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 1 DWDTS-----ALAPYLGT 13
 ||| I ||| |||
 DB 437 DWDDSANERAAFLGT 453

RESULT 8
 E70747
 hypothetical protein Rv2917 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: E70747
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: E70747
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-626 <COL>
 A:Cross-references: GB:274697; GB:AL123456; NID:g3261602; PIDN:CAA98979.1; PID:e326856;
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv2917

Query Match 48.9%; Score 43; DB 2; Length 626;
 Best Local Similarity 61.5%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQ 14
 I: || |||||
 DB 68 WRRALVQYLGTQ 80

RESULT 9
 S15205
 superoxide dismutase (EC 1.15.1.1) (Fe) [validated] - Mycobacterium tuberculosis

C:Species: Mycobacterium tuberculosis
 C:Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 20-Jun-2000
 C:Accession: S15205; C70654; S10908
 R:Zhang, Y.; Lathigra, R.; Garbe, T.; Catty, D.; Young, D.
 Mol. Microbiol. 5, 381-391, 1991
 A:Title: Genetic analysis of superoxide dismutase, the 23 kilodalton antigen of Mycobacterium tuberculosis
 A:Reference number: S15205; MUID:91251768
 A:Accession: S15205
 A:Molecule type: DNA
 A:Residues: 1-207 <ZHA>
 A:Cross-references: EMBL:X52861; NID:g794079; PIDN:CAA37042.1; PID:g581379
 A:Experimental source: strain H37RV
 A:Note: the authors translated the codon GAA for residue 3 as Gln
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: C70654
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-207 <COL>
 A:Cross-references: GB:283864; GB:AL123456; NID:g3261687; PIDN:CAB06220.1; PID:gi7811
 A:Experimental source: strain H37RV
 R:Cooper, J.B.; McIntyre, K.; Wood, S.P.; Zhang, Y.; Young, D.
 submitted to the Brookhaven Protein Data Bank, September 1994
 A:Reference number: A52807; PDB:1IDS
 A:Contents: annotation; X-ray crystallography, 2.0 angstroms, residues 2-199
 R:Cooper, J.B.; Driessen, H.P.C.; Wood, S.P.; Zhang, Y.; Young, D.
 J. Mol. Biol. 235, 1156-1158, 1994
 A:Title: Crystallisation and preliminary X-ray analysis of the iron-dependent superoxo
 A:Reference number: A59029; MUID:94118350
 A:Contents: annotation; X-ray crystallography
 C:Genetics:
 A:Gene: soda
 A:Start codon: GTG
 C:Function:
 A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
 C:Superfamily: superoxide dismutase (Mn)
 C:Keywords: iron; metalloprotein; oxidoreductase; tetramer
 F:28,76,160,164/Binding site: iron (His, His, Asp, His) #status experimental

Query Match 47.7%; Score 42; DB 2; Length 207;
 Best Local Similarity 50.0%; Pred. No. 11;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQ 14
 ||| ||| I:
 DB 10 DWDYGALEPHISGO 23

RESULT 10
 C81748
 DNA gyrase, chain A TC0031 [imported] - Chlamydia muridarum (strain Nigg)
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
 C:Accession: C81748
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe
 Nuclear Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150255
 A:Accession: C81748
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-490 <TET>
 A:Cross-references: GB:AE002271; GB:AE002160; NID:g7190063; PIDN:AAF38922.1; PID:g719
 A:Experimental source: strain Nigg (MoPn)
 C:Genetics:
 A:Gene: TC0031

A:Reference number: S00508; MUID:88122563
A:Accession: S00508
A:Status: not compared with conceptual translation
A:A.Molecule type: mRNA
A:Residues: 1-512,514-564,'Q',565-583,'HV',586-656 <SCH>
A:A.Cross-references: GB:Q66742; NID:9288441; PIDN:CAA29917.1; PID:9288442
A:Accession: S0110
A:Status: not compared with conceptual translation
A:A.Molecule type: DNA
A:Residues: 1-527,'G',528-584,'Q',585-593,'MT',596-599,'LG',602-604,'LRS',605-608,'LRS',609-612,'LRS',613-616,'LRS',617-620,'LRS',621-624,'LRS',625-628,'LRS',629-632,'LRS',633-636,'LRS',637-640,'LRS',641-644,'LRS',645-648,'LRS',649-652,'LRS',653-656,'LRS',657-660,'LRS',661-664,'LRS',665-668,'LRS',669-672,'LRS',673-676,'LRS',677-680,'LRS',681-684,'LRS',685-688,'LRS',689-692,'LRS',693-696,'LRS',697-700,'LRS',701-704,'LRS',705-708,'LRS',709-712,'LRS',713-716,'LRS',717-720,'LRS',721-724,'LRS',725-728,'LRS',729-732,'LRS',733-736,'LRS',737-740,'LRS',741-744,'LRS',745-748,'LRS',749-752,'LRS',753-756,'LRS',757-760,'LRS',761-764,'LRS',765-768,'LRS',769-772,'LRS',773-776,'LRS',777-780,'LRS',781-784,'LRS',785-788,'LRS',789-792,'LRS',793-796,'LRS',797-800,'LRS',801-804,'LRS',805-808,'LRS',809-812,'LRS',813-816,'LRS',817-820,'LRS',821-824,'LRS',825-828,'LRS',829-832,'LRS',833-836,'LRS',837-840,'LRS',841-844,'LRS',845-848,'LRS',849-852,'LRS',853-856,'LRS',857-860,'LRS',861-864,'LRS',865-868,'LRS',869-872,'LRS',873-876,'LRS',877-880,'LRS',881-884,'LRS',885-888,'LRS',889-892,'LRS',893-896,'LRS',897-900,'LRS',901-904,'LRS',905-908,'LRS',909-912,'LRS',913-916,'LRS',917-920,'LRS',921-924,'LRS',925-928,'LRS',929-932,'LRS',933-936,'LRS',937-940,'LRS',941-944,'LRS',945-948,'LRS',949-952,'LRS',953-956,'LRS',957-960,'LRS',961-964,'LRS',965-968,'LRS',969-972,'LRS',973-976,'LRS',977-980,'LRS',981-984,'LRS',985-988,'LRS',989-992,'LRS',993-996,'LRS',997-1000,'LRS',1001-1004,'LRS',1005-1008,'LRS',1009-1012,'LRS',1013-1016,'LRS',1017-1020,'LRS',1021-1024,'LRS',1025-1028,'LRS',1029-1032,'LRS',1033-1036,'LRS',1037-1040,'LRS',1041-1044,'LRS',1045-1048,'LRS',1049-1052,'LRS',1053-1056,'LRS',1057-1060,'LRS',1061-1064,'LRS',1065-1068,'LRS',1069-1072,'LRS',1073-1076,'LRS',1077-1080,'LRS',1081-1084,'LRS',1085-1088,'LRS',1089-1092,'LRS',1093-1096,'LRS',1097-1100,'LRS',1101-1104,'LRS',1105-1108,'LRS',1109-1112,'LRS',1113-1116,'LRS',1117-1120,'LRS',1121-1124,'LRS',1125-1128,'LRS',1129-1132,'LRS',1133-1136,'LRS',1137-1140,'LRS',1141-1144,'LRS',1145-1148,'LRS',1149-1152,'LRS',1153-1156,'LRS',1157-1160,'LRS',1161-1164,'LRS',1165-1168,'LRS',1169-1172,'LRS',1173-1176,'LRS',1177-1180,'LRS',1181-1184,'LRS',1185-1188,'LRS',1189-1192,'LRS',1193-1196,'LRS',1197-1200,'LRS',1201-1204,'LRS',1205-1208,'LRS',1209-1212,'LRS',1213-1216,'LRS',1217-1220,'LRS',1221-1224,'LRS',1225-1228,'LRS',1229-1232,'LRS',1233-1236,'LRS',1237-1240,'LRS',1241-1244,'LRS',1245-1248,'LRS',1249-1252,'LRS',1253-1256,'LRS',1257-1260,'LRS',1261-1264,'LRS',1265-1268,'LRS',1269-1272,'LRS',1273-1276,'LRS',1277-1280,'LRS',1281-1284,'LRS',1285-1288,'LRS',1289-1292,'LRS',1293-1296,'LRS',1297-1300,'LRS',1301-1304,'LRS',1305-1308,'LRS',1309-1312,'LRS',1313-1316,'LRS',1317-1320,'LRS',1321-1324,'LRS',1325-1328,'LRS',1329-1332,'LRS',1333-1336,'LRS',1337-1340,'LRS',1341-1344,'LRS',1345-1348,'LRS',1349-1352,'LRS',1353-1356,'LRS',1357-1360,'LRS',1361-1364,'LRS',1365-1368,'LRS',1369-1372,'LRS',1373-1376,'LRS',1377-1380,'LRS',1381-1384,'LRS',1385-1388,'LRS',1389-1392,'LRS',1393-1396,'LRS',1397-1400,'LRS',1401-1404,'LRS',1405-1408,'LRS',1409-1412,'LRS',1413-1416,'LRS',1417-1420,'LRS',1421-1424,'LRS',1425-1428,'LRS',1429-1432,'LRS',1433-1436,'LRS',1437-1440,'LRS',1441-1444,'LRS',1445-1448,'LRS',1449-1452,'LRS',1453-1456,'LRS',1457-1460,'LRS',1461-1464,'LRS',1465-1468,'LRS',1469-1472,'LRS',1473-1476,'LRS',1477-1480,'LRS',1481-1484,'LRS',1485-1488,'LRS',1489-1492,'LRS',1493-1496,'LRS',1497-1500,'LRS',1501-1504,'LRS',1505-1508,'LRS',1509-1512,'LRS',1513-1516,'LRS',1517-1520,'LRS',1521-1524,'LRS',1525-1528,'LRS',1529-1532,'LRS',1533-1536,'LRS',1537-1540,'LRS',1541-1544,'LRS',1545-1548,'LRS',1549-1552,'LRS',1553-1556,'LRS',1557-1560,'LRS',1561-1564,'LRS',1565-1568,'LRS',1569-1572,'LRS',1573-1576,'LRS',1577-1580,'LRS',1581-1584,'LRS',1585-1588,'LRS',1589-1592,'LRS',1593-1596,'LRS',1597-1600,'LRS',1601-1604,'LRS',1605-1608,'LRS',1609-1612,'LRS',1613-1616,'LRS',1617-1620,'LRS',1621-1624,'LRS',1625-1628,'LRS',1629-1632,'LRS',1633-1636,'LRS',1637-1640,'LRS',1641-1644,'LRS',1645-1648,'LRS',1649-1652,'LRS',1653-1656,'LRS',1657-1660,'LRS',1661-1664,'LRS',1665-1668,'LRS',1669-1672,'LRS',1673-1676,'LRS',1677-1680,'LRS',1681-1684,'LRS',1685-1688,'LRS',1689-1692,'LRS',1693-1696,'LRS',1697-1700,'LRS',1701-1704,'LRS',1705-1708,'LRS',1709-1712,'LRS',1713-1716,'LRS',1717-1720,'LRS',1721-1724,'LRS',1725-1728,'LRS',1729-1732,'LRS',1733-1736,'LRS',1737-1740,'LRS',1741-1744,'LRS',1745-1748,'LRS',1749-1752,'LRS',1753-1756,'LRS',1757-1760,'LRS',1761-1764,'LRS',1765-1768,'LRS',1769-1772,'LRS',1773-1776,'LRS',1777-1780,'LRS',1781-1784,'LRS',1785-1788,'LRS',1789-1792,'LRS',1793-1796,'LRS',1797-1800,'LRS',1801-1804,'LRS',1805-1808,'LRS',1809-1812,'LRS',1813-1816,'LRS',1817-1820,'LRS',1821-1824,'LRS',1825-1828,'LRS',1829-1832,'LRS',1833-1836,'LRS',1837-1840,'LRS',1841-1844,'LRS',1845-1848,'LRS',1849-

C:Species: Chlamydomonas reinhardtii
C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 02-Feb-2001
C:Accession: T08044
R:Wilkinson, C.G.; King, S.M.; Witman, G.B.
J. Cell Sci. 107, 497-506, 1994
A:Title: Molecular analysis of the gamma heavy chain of Chlamydomonas flagellar outer-
A:Reference number: 216311; MUID:94274766
A:Accession: T08044
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4485 <WIL>
A:Cross-references: EMBL:U15303; NID:g557715; PIDN:AAA50455.1; PID:g557716
A:Experimental source: strain 1132D
C:Superfamily: dynein heavy chain, ciliary
C:Keywords: ATP; hydrolase; microtubule binding; nucleotide binding; P-loop
P:1819-1826/Region: nucleotide-binding motif A (P-loop)

Query Match 47.7%; Score 42; DB 2; Length 4485;
Best Local Similarity 46.7%; Pred. No. 3.4e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWTSALAPYLGTQE 15
||: ||: ||: ||
Db 1188 DWANAMVGLDPQE 1202

RESULT 14
E71098
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: E71098
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A:Reference number: A71000; MUID:98344137
A:Accession: E71098
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-217 <KAW>
A:Cross-references: GB:AF000004; NID:g3236131; PIDN:BAA30147.1; PID:g3257464
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1049
C:Superfamily: aroM protein

Query Match 46.6%; Score 41; DB 2; Length 217;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALAPYLGTQEE 16
::: ||: ||: ||
Db 154 SVSPYVGTEEE 164

RESULT 15
DSBYN
superoxide dismutase (EC 1.15.1.1) (Mn) precursor [validated] - yeast (Saccharomyces cer
N:Alternate names: protein YHR008C
C:Species: Saccharomyces cerevisiae
C:Date: 19-Feb-1984 #sequence_revision 17-Mar-1987 #text_change 23-Mar-2001
C:Accession: A00521; S46785; A90766
R:Marres, C.A.M.; Van Loon, A.P.G.M.; Oudshoorn, P.; Van Steeg, H.; Grivell, L.A.; Slate
Eur. J. Biochem. 147, 153-161, 1985
A:Title: Nucleotide sequence analysis of the nuclear gene coding for manganese superoxid
A:Reference number: A91141; MUID:85127011
A:Accession: A00521
A:Molecule type: DNA
A:Residues: 1-233 <MAR>

A:Cross-references: EMBL:X02156; NID:g4513; PIDN:CAA26092.1; PID:g4514
R:Du, Z.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid L2825.
A:Reference number: S46774
A:Accession: S46785
A:Molecule type: DNA
A:Residues: 1-233 <DUZ>
A:Cross-references: EMBL:U10400; NID:g500701; PIDN:AAB68939.1; PID:g500704; GSPDB:GNO
R:ditlow, C.; Johansen, J.T.; Martin, B.M.; Svendsen, I.
Carlsberg Res. Commun. 47, 81-91, 1982
A:Title: The complete amino acid sequence of manganese-superoxide dismutase from Sacc
A:Reference number: A90766
A:Accession: A90766
A:Molecule type: protein
A:Residues: 27-229 <DIR>
C:Genetics:
A:Gene: SGD:SOD2; MIPS:YHR008C
A:Cross-references: SGD:S0001050; MIPS:YHR008C
A:Map position: 8R
A:Genome: nuclear
C:Complex: homotetramer
C:Function:
A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
C:Superfamily: superoxide dismutase (Mn)
C:Keywords: homotetramer; manganese; metalloprotein; mitochondrial matrix; mitochondr
F:1-25/Domain: transit peptide (mitochondrion) #status predicted <TRP>
F:27-229/Product: superoxide dismutase (Mn) #status experimental <WAT>
F:52,107,194,198/Binding site: manganese (His, His, Asp, His) #status predicted

Query Match 46.6%; Score 41; DB 1; Length 233;
Best Local Similarity 53.8%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQ 14
||: ||: ||: ||
Db 35 WDFGALEPIVSGQ 47

RESULT 16
T31269
4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.-) - Sphingomonas aromaticivorans plasmid
C:Species: Sphingomonas aromaticivorans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C:Accession: T31269
R:Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.
submitted to the EMBL Data Library, July 1998
A:Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas arom
A:Reference number: Z20992
A:Accession: T31269
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-343 <ROM>
A:Cross-references: EMBL:AF079317; NID:g3378261; PID:g3378410; PIDN:AAD03993.1
C:Genetics:
A:Gene: xylK
A:Genome: plasmid pNLI
C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 46.6%; Score 41; DB 2; Length 343;
Best Local Similarity 33.3%; Pred. No. 29;
Matches 10; Conservative 2; Mismatches 4; Indels 14; Gaps 1;

QY 1 DWD-----TSALAPYLGTQEE 16
||: ||: ||: ||
Db 70 DWDWIEAAADVKNVLTLLVPGIGTAAE 99

RESULT 17
C69682
gamma-glutamyl phosphate reductase proA - Bacillus subtilis

C:Species: Bacillus subtilis.
 C:Date: 05-Jun-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
 C:Accession: C69682
 R:Thangaraj, H.S.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
 R:Kunst, F.; Ogasawara, N.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 C: Bron, S.; Brouillet, S.; Emmer, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar, E.
 A: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar, E.
 Nature 390, 249-256, 1997.
 A: Authors: Foulger, D.B.; Hertz, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, A.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A: Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetle
 Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
 A: Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowsky, Uchiyama
 T. Winters; Yamahata, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A: Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A: Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A: Reference number: A69580; MUID: 98044033
 A: Accession: C69682
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Residues: 415 <KUN>
 A: Cross-references: GB:299110; GB:AL009126; NID: g2633472; PIDN: CAB13170.1; PID: g2633667
 A: Experimental source: strain 168
 C: Geneticals:
 A: Gene: proA.
 C: Superfamily: glutamate-5-semialdehyde dehydrogenase

Query Match 46.6%; Score 41; DB 2; Length 415;
 Best Local Similarity 50.0%; Pred. No. 36;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 DWDTSALAPYLGTQ 14
 DB 306 DWETFLAPVLSVK 319

RESULT 18
 D72528
 Probable L-asparaginase APE2200 - Aeropyrum pernix (strain K1)
 C: Species: Aeropyrum pernix
 C: Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C: Accession: D72528
 R: Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
 DNA Res. 6, 83-101, 1999
 A: Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
 A: Authors: D72528
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-427 <RAW>
 A: Cross-references: DDBJ:AP000063; NID: g5105654; PIDN: BAA81212.1; PID: g5105900
 A: Experimental source: strain K1
 C: Geneticals:
 A: Gene: APE2200

Query Match 46.6%; Score 41; DB 2; Length 427;
 Best Local Similarity 46.7%; Pred. No. 37;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 1 DWDTSALAPYLGTQF 15
 DB 90 DYETGAVKPYLDASE 104

RESULT 19
 S06599
 Nucleoside diphosphate (EC 1.15.1.1) (Mn) - Mycobacterium leprae
 C: Species: Mycobacterium leprae

C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 18-Jun-1999
 C:Accession: S06599; A37212
 R:Thangaraj, H.S.; Lamb, F.I.; Davis, E.O.; Colston, M.J.
 Nucleic Acids Res. 17, 8378, 1989
 A: Title: Nucleotide and deduced amino acid sequence of Mycobacterium leprae manganese
 A: Reference number: S06599; MUID: 90045970
 A: Molecule type: DNA
 A: Residues: 1-207 <THA>
 A: Cross-references: EMBL: X16453; NID: g44405; PIDN: CAA34472.1; PID: g581343
 R:Thangaraj, H.S.; Lamb, F.I.; Davis, E.O.; Jenner, P.J.; Jeyakumar, L.H.; Colston, M
 Infect. Immun. 58, 1937-1942, 1990
 A: Title: Identification, sequencing, and expression of Mycobacterium leprae superoxid
 A: Accession: A37212
 A: Cross-references: A37212; MUID: 90256282
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 'V', 2-207 <TH2>
 A: Cross-references: GB: X16453
 C: Geneticals:
 C: Function: GTG
 A: Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxyg
 C: Superfamily: superoxide dismutase (Mn)
 A: Keywords: manganese; metalloprotein; oxidoreductase
 F: 28,76,160,164/Binding site: manganese (His, His, Asp, His) #status predicted

Query Match 45.5%; Score 40; DB 2; Length 207;
 Best Local Similarity 54.5%; Pred. No. 24;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 DWDTSALAPYL 11
 DB 10 DWDYALAEPHI 20

RESULT 20
 T31723
 hypothetical protein F40A3.6 - Caenorhabditis elegans
 C: Species: Caenorhabditis elegans
 C: Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
 C: Accession: T31723
 R: Geisel, C.; Bradshaw, H.; Keppler, D.
 submitted to the EMBL Data Library, July 1997
 A: Description: The sequence of C. elegans cosmid F40A3.
 A: Reference number: T31723
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-258 <GEI>
 A: Cross-references: EMBL: AF016423; PIDN: AAB65325.1; GSPDB: GN000023; CESP: F40A3.6
 A: Experimental source: strain Bristol N2; clone F40A3
 C: Geneticals:
 A: Gene: F40A3.6
 A: Molecule type: DNA
 A: Intron: 53/3; 109/3; 175/3; 204/2
 C: Superfamily: Caenorhabditis elegans hypothetical protein F40A3.6

Query Match 45.5%; Score 40; DB 2; Length 258;
 Best Local Similarity 45.6%; Pred. No. 31;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 DWDTSALAPYL 11
 DB 201 DFDSPLSPYL 211

RESULT 21
 PF3371
 hypothetical protein PA2201 [imported] - Pseudomonas aeruginosa (strain PA01)
 C: Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83371
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337
A:Accession: F83371
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <STO>
A:Cross-references: GB:AE004646; GB:AE004091; NID:g9948213; PIDN:AAG05589.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2201

Query Match 45.5%; Score 40; DB 2; Length 294;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQ 14
I | | | | | | |
Db 218 DAQALAPYLQSQ 229

RESULT 22
D85230
Hypothetical 52.9 kD protein in aidB-tpsF intergenic region - *Escherichia coli* (strain K12)
N:Alternate names: hypothetical protein o488
C:Species: *Escherichia coli*
C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: D85230; S56418
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D85230
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-484 <BLAT>
A:Cross-references: GB:AE000491; GB:U00096; NID:g2367357; PIDN:AAC77150.1; PID:g2367358;
A:Experimental source: strain K-12, substrain MG1655
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.0 to 92.9% of the genome
A:Reference number: S56314; MUID:95334362
A:Accession: S56418
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-469, 'AOKMKNNWQNSLLNKEF' <BUR>
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97089.1; PID:g537034
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
C:Genetics:
A:Gene: yJf5
A:Start codon: GTG

Query Match 45.5%; Score 40; DB 2; Length 484;
Best Local Similarity 46.7%; Pred. No. 62;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16
I | | | | | | |
Db 220 WIAYKVAPFLGKKEE 234

RESULT 23
A86116
Hypothetical protein sgat - *Escherichia coli* (strain O157:H7)
C:Species: *Escherichia coli*

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: A86116
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May, M.; Allen, L.; Grothbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamouisis, K.; Apoda, N.; Lory, S.; Olson, M.V.
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A86116
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-484 <STO>
A:Cross-references: GB:AE005174; NID:g12519184; PIDN:AAG59389.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: sgat

Query Match 45.5%; Score 40; DB 2; Length 484;
Best Local Similarity 46.7%; Pred. No. 62;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16
I | | | | | | |
Db 220 WIAYKVAPFLGKKEE 234

RESULT 24
T38093
Probable protein disulfide-isomerase (EC 5.3.4.1) SPAC1F5.02 [similarity] - fission yeast
C:Species: *Schizosaccharomyces pombe*
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
C:Accession: T38093
R:Gentles, S.; Odell, C.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Walsh, S.V.
Submitted to the EMBL Data Library, November 1995
A:Reference number: Z21768
A:Accession: T38093
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-492 <GEN>
A:Cross-references: EMBL:Z68136; PIDN:CAA92230.1; GSPDB:GN00066; SPDB:SPAC1F5.02
A:Experimental source: strain 972h; cosmid c1F5
C:Genetics:
A:Gene: SPDB:SPAC1F5.02
A:Map position: 1
C:Superfamily: protein disulfide-isomerase; thioredoxin homology
C:Keywords: intramolecular oxidoreductase; isomerase

Query Match 45.5%; Score 40; DB 2; Length 492;
Best Local Similarity 41.7%; Pred. No. 64;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLIG 12
I | | | | | | |
Db 215 DWDPSAIADFIG 226

RESULT 25
D86442
Probable amino acid permease [imported] - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D86442
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, U.; Chen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: A86141; MUID:21016719
A:Accession: U88442
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-495 <STO>
A:Cross-references: GB:AE005172; NID:g11136719; PIDN:AAG31300.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 45.5%; Score 40; DB 2; Length 495;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQE 15
DB 116 WVSSALGPWFQ 129

RESULT 26
C84487
hypothetical protein At2g07650 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84487
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Ko, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Nasser, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84487
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-732 <STO>
A:Cross-references: GB:AE002093; NID:g5001454; PIDN:AAD37021.1; GSPDB:GN00139
A:Gene: At2g07650
A:Map position: 2

Query Match 45.5%; Score 40; DB 2; Length 732;
Best Local Similarity 45.5%; Pred. No. 99;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGT 13
DB 500 WDMATIAPYISS 511

RESULT 27
E84547
probable disease resistance protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84547
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
Ko, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Nasser, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84547
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-957 <STO>
A:Cross-references: GB:AE002093; NID:g6598361; PIDN:AAF18599.1; GSPDB:GN00139
A:Gene: At2g17050
A:Map position: 2

Query Match 45.5%; Score 40; DB 2; Length 957;

Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQE 15
DB 335 WDTSIIQPLEDEE 348

RESULT 28
T51314
probable CO-induced hydrogenase complex chain coom [imported] - Rhodospirillum rubrum
C:Species: Rhodospirillum rubrum
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
R:Kerby, R.L.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z25372
A:Accession: T51314
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA; KEB>
A:Cross-references: EMBL:U65510; PIDN:AAC45116.1
A:Experimental source: strain UR1
C:Genetics:
A:Gene: coom

Query Match 45.5%; Score 40; DB 2; Length 1265;
Best Local Similarity 54.5%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGT 12
DB 229 WDAVLLPFG 239

RESULT 29
E84184
hypothetical protein Vng0243c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84184
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Barquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.; Denson, P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genomic analysis of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: E84184
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-575 <STO>
A:Cross-references: GB:AE004437; NID:g10579883; PIDN:AAG18841.1; GSPDB:GN00138
A:Gene: VNG0243C

Query Match 44.9%; Score 39.5; DB 2; Length 575;
Best Local Similarity 50.0%; Pred. No. 92;
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 1 DWDTSALAPYLGTQE 16
DB 338 DWEEAAMA-YLGAPTE 352

RESULT 30
A70825
hypothetical protein rv0750 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70825

R; Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: A70825

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-81 <COL>

A:Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17517.1; PID:el25327

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv0750

Query Match 44.3%; Score 39; DB 2; Length 81;
Best Local Similarity 70.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSALAPYL 11
| | | | |
DB 60 WYTOGLAPYL 69

RESULT 31

G70823

Hypothetical protein Rv0740 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: G70823

R; Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: G70823

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-175 <COL>

A:Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17507.1; PID:el25327

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv0740

Query Match 44.3%; Score 39; DB 2; Length 175;
Best Local Similarity 70.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTSALAPYL 11
| | | | |
DB 154 WYTOGLAPYL 163

RESULT 32

S73413

transcription antitermination factor nusG - Mycoplasma pneumoniae (strain ATCC 29342)

N:Alternate names: hypothetical protein D09_off320

C:Species: Mycoplasma pneumoniae

A:Variety: ATCC 29342

C:Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999

C:Accession: S73413

R; Himmelfreid, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R. Nucleic Acids Res. 24, 4420-4449, 1996

A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S73327; MUID:97105885

A:Accession: S73413

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-320 <HIM>

A:Cross-references: EMBL:AE000011; GB:U00089; NID:g1673740; PIDN:AAB95735.1; PID:g167

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

C:Genetics:

A:Gene: nusG

A:Genetic code: SGC3

Query Match 44.3%; Score 39; DB 2; Length 320;
Best Local Similarity 61.5%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 15
| | | | |
DB 234 DPSVLIPYLGYE 246

RESULT 33

T32099

Hypothetical protein F45C12.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Aug-2000

C:Accession: T32099

R; Johnson, D.; Kramer, J.; Keppler, D. submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid F45C12.

A:Reference number: Z21123

A:Accession: T32099

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-337 <JOH>

A:Cross-references: EMBL:AF016684; PIDN:AAB66207.1; GSPDB:GN00020; CESP:F45C12.8

A:Experimental source: strain Bristol N2; clone F45C12

C:Genetics:

A:Gene: CESP:F45C12.8

A:Map position: 2

A:Introns: 19/2; 50/2; 78/2; 269/3

C:Superfamily: Caenorhabditis elegans hypothetical protein C36C9.3

Query Match 44.3%; Score 39; DB 2; Length 337;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 DWDTLSALAPYLGTQEE 16
| | | | |
DB 286 DMNTSLGLKSVFTQDE 301

RESULT 34

C71103

probable aspartokinase - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999

C:Accession: C71103

R; Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hainaka, Y.; Hino, Y.; Yamamoto, S.; Se M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137

A:Accession: C71103

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-346 <KAW>

A:Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30185.1; PID:d1031128; PID:g32

A:Experimental source: strain OT3

A:Note: This accession replaces an interim accession for a sequence replaced by Genba

C:Genetics:

A:Gene: PH1086

Query Match 44.3%; Score 39; DB 2; Length 346;
Best Local Similarity 77.8%; Pred. No. 64;

C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Jun-1999
C:Accession: A69494
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, R.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343
A:Accession: A69494
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-453 <KLE>
A:Cross-references: GB:AE000968; GB:AE000782; NID:g2689291; PIDN:AAB89301.1; PID:g264858
C:Superfamily: indoleacetamide hydrolase

Query Match 44.3%; Score 39; DB 2; Length 453;
Best Local Similarity 50.0%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 12
: : | | | |
Db 313 EWEKVAFPYLG 324

RESULT 40
H69333
medium-chain acyl-CoA ligase (alkK-3) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 18-Aug-2000
C:Accession: H69333
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, R.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
A:Reference number: A69250; MUID:98049343
A:Accession: H69333
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-504 <KLE>
A:Cross-references: GB:AE001058; GB:AE000782; NID:g2689381; PIDN:AAB90569.1; PID:g264994
C:Superfamily: probable acyl-CoA ligase medium chain; acetate--CoA ligase homology
F:55-497/Domain: acetate--CoA ligase homology <ACL>

Query Match 44.3%; Score 39; DB 2; Length 504;
Best Local Similarity 54.5%; Pred. No. 96;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WDTLSALAPYL 12
: : | | | | | |
Db 212 WGTVFAPYMG 222

RESULT 41
G83339
hypothetical protein PA2448 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83339
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: G83339
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-580 <STO>
A:Cross-references: GB:AE004672; GB:AE004091; NID:g9948492; PIDN:AAG05836.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2448

Query Match 44.3%; Score 39; DB 2; Length 580;
Best Local Similarity 40.0%; Pred. No. 11e+02;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQE 15
: : | | | | :
Db 480 DWPVSSLSPPWKAIQQ 494

RESULT 42
SYECMT
methionine--tRNA ligase (EC 6.1.1.10) [validated] - Escherichia coli
N:Alternate names: methionyl-tRNA synthetase
C:Species: Escherichia coli
C>Date: 13-Jun-1983 #sequence_revision 30-Jun-1992 #text_change 26-May-2000
C:Accession: S14427; A91797; A91123; S11949; A64979; A01187; A35821
R:Dardel, F.
submitted to the EMBL Data Library, October 1990
A:Reference number: S14427
A:Accession: S14427
A:Molecule type: DNA
A:Residues: 1-677 <DARI>
A:Cross-references: EMBL:X55791; NID:g42015; PIDN:CAA39315.1; PID:g42016
R:Dardel, F.; Fayat, G.; Blanquet, S.
J. Bacteriol. 160, 1115-1122, 1984
A:Title: Molecular cloning and primary structure of the Escherichia coli methionyl-tr
A:Reference number: A91797; MUID:85054627
A:Accession: A91797
A:Molecule type: DNA
A:Residues: 2-677 <DAR>
A:Cross-references: GB:K02671; NID:gl46828; PIDN:AAA24161.1; PID:gl46829; GB:J01649;
R:Barker, D.G.; Ebel, J.P.; Jakes, R.; Bruton, C.J.
Eur. J. Biochem. 127, 449-457, 1982
A:Title: Methionyl-tRNA synthetase from E. coli: primary structure of the active crys
A:Reference number: A91123; MUID:83079258
A:Accession: A91123
A:Molecule type: DNA
A:Residues: 2-15, 'V', '17-149, 'L', '151-435, 'A', '437-564 <BAR>
A:Cross-references: GB:K02671; GB:J01649; GB:J01650
A:Note: most of this sequence was confirmed by protein sequencing
R:Dardel, F.; Panvert, M.; Fayat, G.
Mol. Gen. Genet. 223, 121-133, 1990
A:Title: Transcription and regulation of expression of the Escherichia coli methionyl
A:Reference number: S11948; MUID:91080852
A:Accession: S11949
A:Molecule type: DNA
A:Residues: 1-51; 661-677 <DAR>
A:Cross-references: EMBL:X55791
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: A64979
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-677 <BLAT>
A:Cross-references: GB:AE000300; GB:U00096; NID:gl788425; PIDN:AAC75175.1; PID:gl7884
A:Experimental source: Strain K-12, substrain MG1655
R:Zelwer, C.; Risler, J.L.; Brunie, S.
J. Mol. Biol. 155, 63-81, 1982
A:Title: Crystal structure of Escherichia coli methionyl-tRNA synthetase at 2.5 A res
A:Reference number: A92880; MUID:82192427
A:Contents: annotation; X-ray crystallography, 2.5 angstroms
R:Hountondji, C.; Schmitter, J.M.; Beauvallet, C.; Blanquet, S.

```

Query Match          44.3%;      Score 39;  DB 2;  Length 677;
Best Local Similarity 63.6%;      Pred. NO. 1.3e+02;
Matches 7;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

QY      2  WDTSAALPYLG 12
          ||| ||||
Db      230  WDISRDAPYFG 240

```

Search completed: June 28, 2001, 11:57:01
Job time: 242 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:25 ; Search time 22.61 Seconds
(without alignments)
24.241 Million cell updates/sec

Title: US-09-439-313-562

Perfect score: 88

Sequence: 1 DWDTSALAPYLGTQEE 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	46	52.3	215	1	YSCR_SALTY
2	43	48.9	206	1	SODM_MYCAV
3	43	48.9	626	1	YTI7_MYCTU
4	42	47.7	206	1	SODM_MYCLP
5	42	47.7	207	1	SODF_MYCTU
6	42	47.7	643	1	CIK2_DROME
7	42	47.7	656	1	CIK4_DROME
8	42	47.7	4485	1	DYHG_CHLRE
9	41	46.6	233	1	SODM_YEAST
10	41	46.6	415	1	PROA_BACSU
11	40	45.5	206	1	SODM_MYCLE
12	40	45.5	440	1	C2I1_HUMAN
13	40	45.5	484	1	SGAT_ECOLI
14	40	45.5	492	1	PD11_SCHPO
15	39	44.3	320	1	Y054_MYCPN
16	39	44.3	453	1	HCAR_ECOLI
17	39	44.3	453	1	YJ54_ARCFU
18	39	44.3	471	1	DAX1_PIG
19	39	44.3	676	1	SYM_ECOLI
20	39	44.3	682	1	SYM_HAEIN
21	39	44.3	727	1	KGA_RAT
22	39	44.3	730	1	KGA_MOUSE
23	39	44.3	734	1	KGA_PIG
24	39	44.3	735	1	KGA_HUMAN
25	38	43.2	562	1	TPA_HUMAN
26	38	43.2	609	1	UGST_ORYGL
27	38	43.2	609	1	UGST_ORISA
28	38	43.2	767	1	AMTH_SACDI
29	38	43.2	912	1	PCGB_BOVIN
30	38	43.2	2292	1	POLG_EMCVB
31	38	43.2	2292	1	POLG_EMCVD
32	37.5	42.6	260	1	MYX2_CABEL
33	37	42.0	118	1	YNIF_AZOBR

34	37	42.0	199	1	NHAA_PSECL	P27764 pseudomonas
35	37	42.0	206	1	SODM_NOCAS	P53651 nocardia as
36	37	42.0	214	1	DKK1_LACAC	Q59483 lactobacilli
37	37	42.0	283	1	YP97_CABEL	Q09245 caenorhabdi
38	37	42.0	295	1	THTR_RAT	P24329 rattus norv
39	37	42.0	296	1	THTR_BOVIN	P00586 bos taurus
40	37	42.0	296	1	THTR_MOUSE	Q00115 mus musculu
41	37	42.0	302	1	VG03_HSV11	Q00115 ictaluriid h
42	37	42.0	319	1	HPRK_TREPA	Q83600 treponema p
43	37	42.0	358	1	PKX1_HUMAN	P51817 homo sapien
44	37	42.0	379	1	METX_NEIMB	Q91295 neisseria m
45	37	42.0	549	1	AMYG_YEAST	P08019 saccharomyc
46	37	42.0	842	1	LPFC_SALTY	P43662 salmonella
47	37	42.0	993	1	VIA_CMVFN	P17769 cucumber mo
48	37	42.0	993	1	VIA_CMVII	Q83270 cucumber mo
49	37	42.0	993	1	VIA_CMVIX	Q66121 cucumber mo
50	37	42.0	993	1	VIA_CMVNT	Q40976 cucumber mo
51	37	42.0	993	1	VIA_CMVO	P20122 cucumber mo
52	37	42.0	993	1	VIA_CMVY	Q83264 cucumber mo
53	37	42.0	1131	1	PMAL_DUNBI	P54211 dunaliella
54	37	42.0	1290	1	BXCL_CLOBO	P18640 clostridium
55	37	42.0	1447	1	YAO6_CABEL	Q20762 caenorhabdi
56	37	42.0	1459	1	YFIM_CABEL	Q21874 caenorhabdi
57	37	42.0	2331	1	RRPL_MABVM	P31352 marburg vir
58	37	42.0	2331	1	RRPL_MABVP	P35262 marburg vir
59	36.5	41.5	271	1	YAD8_SCHPO	Q09834 schizosacch
60	36.5	41.5	522	1	LEUL_ECOLI	P09151 escherichia
61	36	40.9	33	1	MTMY_MYTED	P81614 mytilus edu
62	36	40.9	153	1	VG59_BPMD2	Q64250 mycobacteri
63	36	40.9	217	1	GTT1_MANSE	P42430 manduca sex
64	36	40.9	256	1	ERG2_NEUCR	Q92254 neurospora
65	36	40.9	262	1	LEC_PHALU	P16300 phaseolus l

ALIGNMENTS

RESULT 1	YSR_SALTY	STANDARD;	PRT;	215 AA.
ID	YSR_SALTY	STANDARD;	PRT;	215 AA.
AC	P74890;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	VIRULENCE PROTEIN YSCR.			
GN	YSR.			
OS	Salmonella typhimurium.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Salmonella.			
OX	NCBI_TaxID=602;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=LT2;			
RA	Hensel M., Shea J.E., Baumber A.J., Gleeson C., Blattner F.R.,			
RA	Holden D.W.;			
RL	Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.			
CC	-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).			
CC	-I- SIMILARITY: BELONGS TO THE FLIP/MOPC/SPAP FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; X99944; CAA68199.1; -			
DR	InterPro; IPR002039; -			
DR	Pfam; PF00813; FLIP; 1.			
DR	PRINTS; PR00951; FLGBIOSNEFLIP.			
DR	PROSITE; PS01060; FLIP_1; 1.			
DR	PROSITE; PS01061; FLIP_2; 1.			

```
KW Virulence; Transmembrane.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
SQ SEQUENCE 215 AA; 24094 MW; 95DDF04BE275A396 CRC64;

Query Match 52.3%; Score 46; DB 1; Length 215;
Best Local Similarity 0.0%; Pre: 1; 1; Indels 0; Gaps 0;
Matches 7; Conservative 2; Mismatches 0;

QY 1 DWDTSALAPY 10
DB 92 EWDKALAPY 101

RESULT 2
SODM MYCAV STANDARD; PRT; 206 AA.
AC P47201;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 39, Last sequence update)
DE HYPOTHETICAL 68.4 KDA PROTEIN RV2917.
GN SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
OS SODA OR SOD.
OS Mycobacterium avium.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1764;
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN=724;
RA MEDLINE=96276149; PubMed=8692009;
RA Escuyer V.E., Haddad N., Frehel C., Berche P.;
RT "Molecular characterization of a surface-exposed superoxide dismutase
of Mycobacterium avium."
RA Microb Drug Resist 2:441-455(1996).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) -> O(2) + H(2)O(2).
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: U11550; AAB08770.1;
CC HSP: P17670; LIDS.
CC InterPro: IPR001189;
CC Pfam: PF00081; SOD_MN: 1.
CC PROSITE: PS00088; Manganese.
CC Oxidoreductase; Manganese.
CC METAL 27 27 BY SIMILARITY.
CC METAL 75 75 MANGANESE (BY SIMILARITY).
CC METAL 159 159 MANGANESE (BY SIMILARITY).
CC METAL 163 163 MANGANESE (BY SIMILARITY).
CC METAL 163 163 MANGANESE (BY SIMILARITY).
CC SEQUENCE 206 AA; 22912 MW; D6B49083AEBF3C98 CRC64;

Query Match 48.9%; Score 43; DB 1; Length 206;
Best Local Similarity 50.0%; Pred. NO. 4.1;
Matches 7; Conservative 3; Mismatches 0; Gaps 0;

QY 1 DWDTSALAPYLGTO 14
DB 9 DWDYAALEPHISGO 22

us-09-439-313-562.rsp
```

```
RESULT 3
YT17 MYCTU STANDARD; PRT; 626 AA.
ID YT17 MYCTU
AC Q10966; Q10967.1; 34, Created)
DT 01-OCT-1996 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 68.4 KDA PROTEIN RV2917.
GN RV2917 OR MTCY338.05.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brisch R., Karkhanavala V., Barry C.E. III, Tekala F.,
RA Bock C.K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Sultson J.E.,
RA Taylor K., Whitehead S., Wren S., Young I.P., Barrell A.N.;
RA "Genomic and biological study of Mycobacterium tuberculosis from the
complete genome sequence."
RA Nature 393:537-544(1998).
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: Z74697; CAA98979.1;
CC Tuberculist; RV2917;
CC NCBI_TaxID=1773;
CC SEQUENCE 626 AA; 68365 MW; C8953B2E52B0B841 CRC64;

Query Match 48.9%; Score 43; DB 1; Length 626;
Best Local Similarity 61.5%; Pred. NO. 12;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WPTSALAPYLGTO 14
DB 68 WRRALVQVYLGTO 80

RESULT 4
SODM MYCLP STANDARD; PRT; 206 AA.
AC O86165;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
OS Mycobacterium lepraemurium.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=64667;
RN RP
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Nakamura M.
RT "Molecular characterization of a surface-exposed superoxide dismutase
of Mycobacterium lepraemurium."
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
```

CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
 CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).
 CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D13288; BAA28850.1; -
 DR HSSP: P17670; 1IDS.
 DR InterPro: IPR001189; -
 DR Pfam: PF00081; sodfe; 1.
 DR PROSITE: PS00088; SOD_MN; 1.
 KW Oxidoreductase; Manganese.
 FT INIT MET 0 0 BY SIMILARITY.
 FT METAL 27 27 MANGANESE (BY SIMILARITY).
 FT METAL 75 75 MANGANESE (BY SIMILARITY).
 FT METAL 159 159 MANGANESE (BY SIMILARITY).
 FT METAL 163 163 MANGANESE (BY SIMILARITY).
 SQ SEQUENCE 206 AA; 23031 MW; 92A063471FA9F22F CRC64;
 Query Match 47.7%; Score 42; DB 1; Length 206;
 Best Local Similarity 50.0%; Pred. No. 6;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 DWDTSALAPYLGTQ 14
 Db 9 DWDYEALPHISGQ 22
 RESULT 5
 SODE MYCTU STANDARD; PRT; 207 AA.
 AC P17670; P96231;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SUPEROXIDE DISMUTASE [FE] (EC 1.15.1.1)
 GN SODE OR SODA OR SOD OR RV3846 OR MTCY01A6.22C.
 OS Mycobacterium tuberculosis, and Mycobacterium bovis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=ERDMANN;
 RX MEDLINE=91251768; PubMed=1904126;
 RA Zhang Y.;
 RT "Genetic analysis of superoxide dismutase, the 23 kilodalton antigen
 RT of Mycobacterium tuberculosis";
 RL Mol. Microbiol. 5:381-391(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RL Nature 393:537-544(1998).
 RN [3]

RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=ERDMANN;
 RX MEDLINE=99134360; PubMed=9933629;
 RA Harth G., Horwitz M.A.;
 RT "Export of recombinant Mycobacterium tuberculosis superoxide dismutase
 RT is dependent upon both information in the protein and mycobacterial
 RT export machinery. A model for studying export of leaderless proteins
 RT by pathogenic mycobacteria";
 RL J. Biol. Chem. 274:4281-4292(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=BCG / PASTEUR;
 RA Kimble E., Sanderson R.J., Gill R.E.;
 RT "Superoxide dismutase of M. bovis BCG";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RC SPECIES=M.tuberculosis;
 RX MEDLINE=95182461; PubMed=7877174;
 RA Cooper J.B., McIntyre K., Badasso M.O., Wood S.P., Zhang Y.,
 RA Garbe T.R., Young D.;
 RT "X-ray structure analysis of the iron-dependent superoxide dismutase
 RT from Mycobacterium tuberculosis at 2.0-A resolution reveals novel
 RT dimer-dimer interactions";
 RL J. Mol. Biol. 246:531-544(1995).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF MUTANT ALA-152.
 RC SPECIES=M.tuberculosis;
 RX MEDLINE=96244503; PubMed=8674528;
 RA Cooper J.B., Seward S., Erskine P.T., Badasso M.O., Wood S.P.,
 RA Zhang Y., Young D.;
 RT "X-ray structure analysis of an engineered Fe-superoxide dismutase
 RT Gly-Ala mutant with significantly reduced stability to denaturant.";
 RL FEBS Lett. 387:105-108(1996).
 CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
 CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
 CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) + H(2)O(2).
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
 CC FAMILY.
 CC -1- CAUTION: ALTHOUGH FOUND EXTRACELLULARLY, NO SIGNAL SEQUENCE IS
 CC PRESENT. AN ALTERNATIVE SECRETORY PATHWAY MAY BE USED.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52861; CAA37042.1; -
 DR EMBL: Z83864; CAB06220.1; -
 DR EMBL: AF061030; AAD15824.1; -
 DR EMBL: AF077406; AAC27527.1; -
 DR FIR: S10908; S10908.
 DR PIR: S15205; S15205.
 DR PDB: 1DS; 20-DEC-94.
 DR Tuberculist: RV3846; -
 DR InterPro: IPR001189; -
 DR Pfam: PF00081; sodfe; 1.
 DR PROSITE: PS00088; SOD_MN; 1.
 KW Oxidoreductase; Iron; 3D-structure.
 FT METAL 28 28 IRON.
 FT METAL 76 76 IRON.
 FT METAL 160 160 IRON.
 FT METAL 164 164 IRON.
 SQ SEQUENCE 207 AA; 23034 MW; DEE9F5921DABE54A CRC64;
 Query Match 47.7%; Score 42; DB 1; Length 207;
 Best Local Similarity 50.0%; Pred. No. 6;

FT	DOMAIN	1699	1727	COILED COIL (POTENTIAL).
FT	DOMAIN	1917	1945	MICROTUBULE-BINDING (POTENTIAL).
FT	DOMAIN	3077	3099	COILED COIL (POTENTIAL).
FT	DOMAIN	3196	3227	COILED COIL (POTENTIAL).
FT	DOMAIN	3265	3343	COILED COIL (POTENTIAL).
FT	DOMAIN	3569	3663	COILED COIL (POTENTIAL).
FT	NP_BIND	1819	1826	ATP (POTENTIAL).
FT	NP_BIND	2099	2106	ATP (POTENTIAL).
FT	NP_BIND	2425	2432	ATP (POTENTIAL).
FT	NP_BIND	2802	2809	ATP (POTENTIAL).
SO	SEQUENCE	4485 AA;	512836 MW;	974B79328B403677 CRC64;

Query Match 47.7%; Score 42; DB 1; Length 4485;
 Best Local Similarity 46.7%; Pred. No. 1.2e+02;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DWDTSALAPYLGTQE 15
 ||: ||: ||: ||: ||

Db 1188 DWEANAMVPLDPOE 1202

RESULT 9

ID	SODM_YEAST	STANDARD;	PRT;	233 AA.
AC	P00447;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	23-OCT-1986 (Rel. 02, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	SUPEROXIDE DISMUTASE [MN], MITOCHONDRIAL PRECURSOR (EC 1.15.1.1).			
GN	SOD2 OR YHR008C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=85137011; PubMed=3882422;			
RA	Marres C.A.M., van Loon A.P.G.M., Oudshoorn P., van Steeg H.,			
RA	Grivell L.A., Slater E.C.;			
RA	"Nucleotide sequence analysis of the nuclear gene coding for			
RT	manganese superoxide dismutase of yeast mitochondria, a gene			
RT	previously assumed to code for the Rieske iron-sulphur protein.";			
RT	Eur. J. Biochem. 147:153-161(1985).			
RN	[2]			
RN	SEQUENCE FROM N.A.			
RC	STRAIN=S288C / AB972;			
RX	MEDLINE=94378003; PubMed=8091229;			
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,			
RA	Kucaba, Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,			
RA	Du Z., Teller L., Hillier L., Jier M., Johnston L., Langston Y.,			
RA	Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,			
RA	Nigmati D., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,			
RA	Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,			
RA	Vaudin M.;			
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome			
RT	VIII.";			
RN	[3]			
RN	Science 265:2077-2082(1994).			
RX	SEQUENCE OF 1-39 FROM N.A.			
RX	MEDLINE=89211942; PubMed=3072251;			
RA	Schrank I.S., Sims P.F., Oliver S.G.;			
RT	"Functional expression of the yeast Mn-superoxide dismutase gene in			
RT	Escherichia coli requires deletion of the signal peptide sequence.";			
RN	Gene 73:121-130(1988).			
RN	[4]			
RN	SEQUENCE OF 27-233.			
RA	Ditlow C., Johansen J.T., Martin B.M., Svendsen I.;			
RT	"The complete amino acid sequence of manganese-superoxide dismutase			
RT	from Saccharomyces cerevisiae.";			
RL	Carlsberg Res. Commun. 47:81-91(1982).			
CC	-1- FUNCTION: DESPOYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE			
CC	CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.			

 NADP(+) = L-GAMMA-GLUTAMYL 5-PHOSPHATE + NADPH.
 -1- PATHWAY: SECOND STEP IN PROLINE BIOSYNTHESIS PATHWAY.
 -1- SIMILARITY: BELONGS TO THE GAMMA-GLUTAMYL PHOSPHATE REDUCTASE
 FAMILY.

 THIS SWISS-PRO entry is copyrighted, it is produced through a collaboration
 between the European Bioinformatics Institute and the EMBL outstation in
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement>
 or send an email to license@isb-sib.ch).

 EMBL: D26044: BAA05045.1
 EMBL: A9002471: CAA05592.1
 EMBL: J29110: CAB13170.1
 Subtilist: BG10964: proA.
 InterPro: IPR000965;
 InterPro: IPR002086;
 Pfam: PF00171: aldedh; 1.
 PROSITE: PS01223; PROA; 1.
 Conflicts: 108
 CONFLICT 174
 CONFLICT 174
 CONFLICT 271
 CONFLICT 359
 CONFLICT 374
 E -> FGFYGAEGIGSTQKLHARGMGLPALSTKYIKGTGQIR
 E -> FGIIRLNRRHQAACKNDGASCTDFYKRIH
 (IN REF.)
 SEQUENCE 415 AA; 45336 MW; 8CA4B0D35F9F62D0 CRC64;

 Query Match 46.6%; Score 41; DB 1; Length 415;
 Best Local Similarity 50.0%; Pred. NO. 17;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0

 QY 1 DWDTSALAPVLCGTQ 14
 DDD 306 DWETEFAPLVSVK 319

 RESULT 11
 ID SODM_MYCLE STANDARD; PRT; 206 AA.
 ID SODM_MYCLE
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 P1367; 1990 (Rel. 13, Created)
 01-OCT-1996 (Rel. 34, Last sequence update)
 01-OCT-1996 (Rel. 34, Last annotation update)
 SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
 SODA.
 OS
 OC Mycobacterium leprae.
 OC Bacteria; Firmicutes

```

CC CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X16453; CAA34472.1; -
CC PIR; S06599; S06599.
CC PIR; A37212; A37212.
CC HSSP; P17670; 1IDS.
CC InterPro; IPR001189; -.
CC Pfam; PF00081; sodfe; 1.
CC PROSITE; PS00088; SOD_MN; 1.
CC Oxidoreductase; Manganese.
CC INIT_MET 0 0 BY SIMILARITY.
CC METAL 27 27 MANGANESE (BY SIMILARITY).
CC METAL 75 75 MANGANESE (BY SIMILARITY).
CC METAL 159 159 MANGANESE (BY SIMILARITY).
CC METAL 163 163 MANGANESE (BY SIMILARITY).
CC SEQUENCE 206 AA; 23027 MW; B02B96EB433F6FF0 CRC64;
SQ
Query Match 45.5%; Score 40; DB 1; Length 206;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 11
DB 9 DWDYAALEPHI 19

RESULT 12
C21L_HUMAN STANDARD; PRT; 440 AA.
AC Q9NVD3;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE PUTATIVE PROTEIN C21ORF18.
GN C21ORF18.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahari K., Masuho Y., Sasaki N.;
RA "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO S.POMBE C1223.04C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AK001660; BAA01819.1; -
CC Hypothetical protein.
CC SEQUENCE 440 AA; 50416 MW; 9EBCAA05397BE287 CRC64;
SQ
Query Match 45.5%; Score 40; DB 1; Length 440;
Best Local Similarity 88.9%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 DTSALAPYL 11
DB 224 DTCALAPYL 232

RESULT 13
SGAT_ECOLI STANDARD; PRT; 484 AA.
AC P39301;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PUTATIVE TRANSPORT PROTEIN SGAT.
GN SGAT.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MGI655;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN=K12 / MGI655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP DISCUSSION OF SEQUENCE.
RA Reizer J., Charbit A., Reizer A., Saier M.H. Jr.;
RT "Novel phosphotransferases system genes revealed by bacterial genome
RT analysis: operons encoding homologues of sugar-specific permease
RT domains of the phosphotransferase system and pentose catabolic
RT enzymes.";
RL Genome Sci. Technol. 1:53-75(1996).
CC -!- FUNCTION: COULD ACT AS THE TRANSPORT PROTEIN FOR THE UNKNOWN
CC PENTITOL SUBSTRATE OF THE SGA OPERON.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: TO M.PNEUMONIAE SGAT HOMOLOG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14003; AAA97089.1; -
CC EMBL; AF000491; AAC77150.1; -
CC EcoGene; EG12493; sgat.
CC Transport; Transmembrane; Inner membrane.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 282 302 POTENTIAL.
FT TRANSMEM 335 355 POTENTIAL.

```

FT TRANSMEM 357 377 POTENTIAL.
FT TRANSMEM 398 418 POTENTIAL.
FT TRANSMEM 420 440 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
FT CONFLICT 470 484 RAEDAEKOLAEQSA -> AOKMKNNWNSLLNREF
(IN REF. 1).
SQ SEQUENCE 484 AA; 52950 MW; 2CBEDD0044BC6CAC CRC64;

Query Match 45.5%; Score 40; DB 1; Length 484;
Best Local Similarity 46.7%; Pred. No. 30;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 WTSALAPYLGTQRE 16
DB 220 WIAYKVAFLQKKEE 234

RESULT 14
PDII_SCHPO STANDARD; PRT; 492 AA.
ID PDII_SCHPO
AC PDII_SCHPO
DT 01-FEB-1996 (Rel. 33, Last Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PUTATIVE PROTEIN DISULFIDE ISOMERASE C1F5.02
GN SPAC1F5.02
OS Schizosaccharomyces pombe (Fission Yeast)
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetes; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=92; Odell C., Churcher C.M., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RA Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PARTICIPATES IN THE FOLDING OF PROTEINS CONTAINING
CC DISULFIDE BONDS. MAY BE INVOLVED IN GLYCOSYLATION, PROLYL
CC HYDROXYLATION AND TRIGLYCERIDE TRANSFER (BY SIMILARITY).
CC -1- DISULFIDE BONDS IN PROTEINS TO FORM THE NATIVE STRUCTURES
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 2 THIOREDOXIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z68136; CAA92230.1;
CC HSP; P07237; IMEK;
CC InterPro; IPR000886;
CC Pfam; PF00085; Thired; 2.
CC PRINTS; PR00421; THIOREDIXIN.
CC PROSITE; PS00014; ER_TARGET; 1.
CC PROSITE; PS00194; THIOREDIXIN; 2.
CC Hypothetical protein; Redox-active center; Isomerase;
CC Endoplasmic reticulum; Repeat; Potential.
CC CHAIN 23 492 PUTATIVE PROTEIN DISULFIDE ISOMERASE
FT CHAIN 23 492 C1F5.02.
FT FT 51 54 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 385 388 REDOX-ACTIVE (BY SIMILARITY).
FT FT 161 161 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 439 492 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 439 492 PRESENT SUGGESTION FROM ER (POTENTIAL).
SQ SEQUENCE 492 AA; 54880 MW; C30B59236566667 CRC64;

Query Match 44.3%; Score 39; DB 1; Length 320;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQRE 15
DB 234 DPSVLPIYLKYE 246

RESULT 16
HCAE_ECOLI STANDARD; PRT; 453 AA.
ID HCAE_ECOLI
AC Q47139; P77590;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-2000 (Rel. 40, Last annotation update)
DE 3-PHENYLPROPIONATE DIOXYGENASE ALPHA SUBUNIT (EC 1.14.1.-)
DE (DIOXYGENIN ALPHA SUBUNIT).
GN HCAE OR PHD1 OR HCA1 OR HCA1 OR DIGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia coli.
OC Escherichia coli.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 45.5%; Score 40; DB 1; Length 492;
Best Local Similarity 41.7%; Pred. No. 30;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 DTSALAPYLK 12
DB 215 DWDPSIADDFIG 226

RESULT 15
Y054_MYCPN STANDARD; PRT; 320 AA.
ID Y054_MYCPN
AC Y054_MYCPN
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHELICAL PROTEIN MG054 HOMOLOG (D09_ORF320).
GN MPN067 OR MP087.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasma pneumoniae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129; 8948633;
RA HELDLINE 9708585;
RA Hermann R.; Hilbert H.; Piagens H.; Pirk E.; Li B.-C.;
RA "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SIMILARITY: BELONGS TO THE NUSG FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000011; AAB95735.1;
CC Hypothetical protein.
SQ SEQUENCE 320 AA; 36107 MW; ECE3CG26C8EAE63F CRC64;


```

Best Local Similarity 64.3%; Pred. No. 41;
Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 1 DWDTSA-LAPYLIG 12
   :||||| | |||
DB 155 NWDTSAPGLRDYIG 168

RESULT 17
YJ54_ARCFU
ID YJ54_ARCFU STANDARD; PRT; 453 AA.
AC O28325;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE AMIDASE AF1954 (EC 3.5.1.4).
GN AF1954.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
EX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.D., Kervagade A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.L., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -1- CATALYTIC ACTIVITY: A MONOCARBOXYLIC ACID AMIDE + H(2)O =
CC -1- A MONOCARBOXYLATE + NH(3).
CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY. STRONG, TO GATA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; AF000968; AAB89301.1; -.
CC TIGR; AF1954; -.
DR InterPro: IPR000120;
DR Pfam: PF01425; Amidase. 1.
DR PROSITE; PS00571; AMIDASES; 1.
KW Hypothetical protein; Hydrolase.
SQ SEQUENCE 453 AA; 50062 MW; 72E8D6A2A535F483 CRC64;

Query Match 44.3%; Score 39; DB 1; Length 453;
Best Local Similarity 50.0%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLIG 12
   :|: | ||||
DB 313 EWKVAFPYPIG 324

RESULT 18

```

ID	DAXI_1
AC	P79386

DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ORPHAN NUCLEAR RECEPTOR DAX-1
 GN NR0B1 OR ARHG OR DAX1
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 CC [1] TaxID=9823;
 CC [2] TISSUE=Adrenal gland;
 CC MEDLINE=98113985; PubMed=9453240;
 CC Parma P., Pailhoux E., Puissant C., Cotinot C.;
 CC "The Dax-1 gene: isolation and expression during gonadal
 CC development"; Mol. Cell. Endocrinol. 135:49-58(1997).
 CC [2] SEQUENCE OF 1-390 FROM N.A.
 CC Behdjani R., Silversides D.W.;
 CC Submitted (Sep1997) to the EMBL/GenBank/DBJ databases.
 CC FOR DEVELOPMENT OF STEROIDGENIC TISSUES ACTS AS A DOMINANT
 CC NEGATIVE REGULATOR OF TRANSCRIPTION MEDIATED BY THE RETINOIC ACID
 CC RECEPTOR (BY SIMILARITY).
 CC [-] SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC [-] SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 CC NO SUBFAMILY. LACKS DNA-BINDING REGION.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC --send an email to: license@isb-sib.ch).
 CC EMBL; U82466; AAB81101.1;
 CC EMBL; AF019044; AAB70254.1;
 CC InterPro; IPR000536; -;
 CC Pfam; PF00104; hormone_rec; 1.
 CC Receptor; Nuclear protein; Transcription regulation; Repressor;
 CC DOMAIN 1 253 4 X 67 AA TANDEM REPEATS.
 CC REPEAT 1 67 1.
 CC REPEAT 68 134 2.
 CC REPEAT 135 201 3.
 CC REPEAT 202 253 4 (INCOMPLETE). (BY SIMILARITY).
 CC REPEAT 124 125 GR -> AG (IN REF. 2).
 CC CONFLICT 243 244 AO -> OR (IN REF. 2).
 CC CONFLICT 266 266 L -> V (IN REF. 2).
 CC CONFLICT 289 289 S -> T (IN REF. 2).
 CC SEQUENCE 471 AA; 52128 MW; 8EE7D133A4677950 CRC64;
 Query Match Similarity 44.38; Score 39; DB 1; Length 471;
 Best Local Similarity 53.38; Pred. No. 42;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 OY 2 WPTSALAPYLCTQEE 16
 DB 173 WDRSYCAQRLGAREE 187
 RESULT 19
 ID SYM_ECOLI STANDARD; PRT; 676 AA.
 AC P00959; 1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)

DE (METERS).
 GN Escherichia coli.
 OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 CC [1] TaxID=562;
 CC [2] SEQUENCE FROM N.A. PubMed=6094501;
 CC Dardel F., Fayat G., Blanquet S.;
 CC "Molecular cloning and primary structure of the Escherichia coli
 CC methionyl-tRNA synthetase gene."; J. Bacteriol. 160:1115-1122(1984).
 CC [2] SEQUENCE FROM N.A.
 CC MEDLINE=91080852; PubMed=2259334;
 CC Dardel F., Panvert M., Fayat G.;
 CC "Transcription and regulation of expression of the Escherichia coli
 CC methionyl-tRNA synthetase gene."; Mol. Gen. Genet. 223:121-133(1990).
 CC [2] SEQUENCE FROM N.A.
 CC STRAIN=K12; BHB2600;
 CC Richters P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
 CC Church G.M.;
 CC Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 CC [4] SEQUENCE FROM N.A.
 CC STRAIN=K12; MG1655;
 CC MEDLINE=97426617; PubMed=9278503;
 CC Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 CC Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 CC Watson B., Shaoh S.;
 CC "Complete genome sequence of Escherichia coli K-12."; Science 277:1453-1474(1997).
 CC [5] SEQUENCE OF 1-565 FROM N.A., AND PARTIAL SEQUENCE.
 CC MEDLINE=83079258; PubMed=6756915;
 CC Barker D.G., Ebel J.-P., Jakes R., Bruton C.J.;
 CC "Methionyl-tRNA synthetase from Escherichia coli. Primary structure
 CC of the amino-terminal fragment."; Eur. J. Biochem. 127:449-457(1982).
 CC [6] X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 CC Zelwer C., Risler J.L., Brunie S.;
 CC "Crystal structure of Escherichia coli methionyl-tRNA synthetase at
 CC 2.5 Å resolution."; J. Mol. Biol. 155:63-81(1982).
 CC [7] X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 CC Brunie S., Zelwer C., Risler J.L.;
 CC "Resolution of the interaction of
 CC methionyl-tRNA synthetase from Escherichia coli with ATP."; J. Mol. Biol. 216:411-424(1990).
 CC [8] X-RAY CRYSTALLOGRAPHY (2.03 ANGSTROMS) OF 1-552.
 CC MEDLINE=20069948; PubMed=10600385;
 CC Mechulam T., Schmitt E., Maveyraud L., Zelwer C., Nureki O.,
 CC "Crystal structure of Escherichia coli methionyl-tRNA synthetase
 CC highlights species-specific features."; J. Mol. Biol. 294:1287-1297(1999).
 CC [9] STRUCTURE BY NMR OF 138-163.
 CC MEDLINE=93294839; PubMed=8515466;
 CC "Methionyl-tRNA synthetase zinc binding domain. Three-dimensional
 CC structure and homology with rubredoxin and gag retroviral proteins."; J. Mol. Biol. 231:1078-1089(1993).

[10]
RP ACTIVE SITE MAPPING.
RX MEDLINE=91084494; PubMed=1702021;
RA Hountondji C., Schmitter J.-M., Beauvallet C., Blanquet S.;
RT "Mapping of the active site of *Escherichia coli* methionyl-tRNA
synthetase: identification of amino acid residues labeled by
periodate-oxidized tRNA(fmet) molecules having modified lengths at
the 3'-acceptor end.";
RL Biochemistry 29:8190-8198(1990).
[11]
RN MUTAGENESIS.
RP MEDLINE=92070503; PubMed=1959615;
RX Fourmy D., Mechulam Y., Brunie S., Blanquet S., Fayat G.;
RA "Identification of residues involved in the binding of methionine by
RT *Escherichia coli* methionyl-tRNA synthetase.";
RL FEBS Lett. 292:259-263(1991).
[12]
RN MUTAGENESIS OF ZINC-LINGANDS.
RP MEDLINE=93294858; PubMed=8515465;
RX Fourmy D., Melnell T., Mechulam Y., Blanquet S.;
RA "Mapping of the zinc binding domain of *Escherichia coli*
RT methionyl-tRNA synthetase.";
RL J. Mol. Biol. 231:1068-1077(1993).
[13]
RN REVIEW.
RP MEDLINE=91129305; PubMed=2126467;
RX Melnell T., Mechulam Y., Dardel F., Schmitter J.M., Hountondji C.,
RA Brunie S., Dessen P., Fayat G., Blanquet S.;
RT "Methionyl-tRNA synthetase from *E. coli* -- a review.";
RL Biochimie 72:625-632(1990).
CC -!- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
CC FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
CC tRNA(fMET) AMINOACYLATION.
CC -!- CATALYTIC ACTIVITY: ATP + L-METHIONINE + tRNA(MET) = AMP +
CC PYROPHOSPHATE + L-METHIONYL-tRNA(MET).
CC -!- COFACTOR: BINDS ONE ZINC ION.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-tRNA SYNTHETASE FAMILY.
CC STRONG, TO CYSTEINYL-tRNA SYNTHETASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; K02671; AA24161.1; -
CC EMBL; X5791; CA39315.1; -
CC EMBL; U00007; AA60526.1; ALT_INIT.
CC EMBL; AE000300; AAC75175.1; -
CC PIR; S14427; SYECMT.
CC PIR; A35821; A35821.
CC PIR; S11949; S11949.
CC PDB; 1MEA; 31-JAN-94.
CC PDB; 1MED; 31-JAN-94.
CC PDB; 1OQT; 01-JAN-00.
CC PDB; 1OQT; 01-JAN-00.
CC SWISS-2DPAGE; P00959; COLI.
CC ECODBASE; F072.0; 6TH EDITION.
CC ECODBASE; G072.0; 6TH EDITION.
CC EcoGene; EG10586; metG.
CC InterPro; IPR001412; -
CC InterPro; IPR002300; -
CC InterPro; IPR002304; -
CC InterPro; IPR002547; -
CC Pfam; PF00133; tRNA-synt_1; 1.
CC Pfam; PF01588; tRNA_bind; 1.
CC PRINTS; PR01041; TRNASYNTHET.
CC PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

KW tRNA-binding: Metal-binding; zinc; 3D-structure.
FT INIT_MET 0
FT SITE 14 24 "HIGH" REGION.
FT SITE 332 336 "KMSKS" REGION.
FT DOMAIN 580 674 tRNA BINDING.
FT METAL 145 145 ZINC.
FT METAL 148 148 ZINC.
FT METAL 158 158 ZINC.
FT METAL 161 161 ZINC.
FT BINDING 335 335 ATP.
FT MUTAGEN 335 335 K->Q,A,E,R: LOSS OF ACTIVITY.
FT STRAND 141 141
FT TURN 142 143
FT STRAND 144 144
FT TURN 146 148
SQ SEQUENCE 676 AA; 76123 MW; 917CD98FE70E7728 CRC64;

Query Match 44.3%; Score 39; DB 1; Length 676;
Best Local Similarity 63.6%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDSALAPYLG 12
||| |||
DB 229 WDISRDAPYFG 239

RESULT 20
SYM_HAEIN
ID SYM_HAEIN STANDARD; PRT; 682 AA.
AC P43828;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE METHIONYL-tRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--tRNA LIGASE)
DE (METRS).
DE METG OR H1276.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
CC FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
CC tRNA(fMET) AMINOACYLATION.
CC -!- CATALYTIC ACTIVITY: ATP + L-METHIONINE + tRNA(MET) = AMP +
CC PYROPHOSPHATE + L-METHIONYL-tRNA(MET).
CC -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-tRNA SYNTHETASE FAMILY.
CC STRONG, TO CYSTEINYL-tRNA SYNTHETASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

RA MEDLINE=88054470; PubMed=2824147;
RA Reddy V.B., Garramone A.J., Sasak H., Wei C.-M., Watkins P., Galli J.,
RA Hsiung N.;
RT "Expression of human uterine tissue-type plasminogen activator in
RT mouse cells using BPV vectors.";
RL DNA 6:461-472(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=86196143; PubMed=3009482;
RA Frieznar Degen S.J., Rajput B., Reich E.;
RT "The human tissue plasminogen activator gene";
RL J. Biol. Chem. 261:6972-6985(1986).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=84298137; PubMed=6089198;
RA Ny T., Elgh F., Lund B.;
RT "The structure of the human tissue-type plasminogen activator gene:
RT correlation of intron and exon structures to functional and
RT structural domains";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5355-5359(1984).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=86284200; PubMed=3090401;
RA Harris T.J., Patel T., Marston F.A., Little S., Entage J.S.,
RA Opdenakker G., Volckaert G., Rombauts W., Billiau A., Somer P.;
RT "Cloning of cDNA coding for human tissue-type plasminogen activator
RT and its expression in Escherichia coli.";
RL Mol. Biol. Med. 3:279-292(1986).
RN [7]
RP SEQUENCE OF 212-361 FROM N.A.
RX MEDLINE=83169656; PubMed=6572897;
RA Edlund T., Ny T., Raanby M., Heden L.-O., Palm G., Holmgren E.,
RA Josephson S.;
RT "Isolation of cDNA sequences coding for a part of human tissue
RT plasminogen activator";
RL Proc. Natl. Acad. Sci. U.S.A. 80:349-352(1983).
RN [8]
RP SEQUENCE OF 1-36 FROM N.A.
RX MEDLINE=85289338; PubMed=3161893;
RA Fisher R., Waller E.K., Grossi G., Thompson D., Tizard R.,
RA Schleuning W.-D.;
RT "Isolation and characterization of the human tissue-type plasminogen
RT activator structural gene including its 5' flanking region.";
RL J. Biol. Chem. 260:11233-11230(1985).
RN [9]
RP SEQUENCE OF 31-562 FROM N.A.
RX MEDLINE=91291340; PubMed=1368681;
RA Itagaki Y., Yasuda H., Morinaga T., Mitsuda S., Higashio K.;
RT "Purification and characterization of tissue plasminogen activator
RT secreted by human embryonic lung diploid fibroblasts, IMR-90 cells.";
RL Agric. Biol. Chem. 55:1225-1232(1991).
RN [10]
RP SEQUENCE OF 36-562.
RC TISSUE-Melanoma;
RX MEDLINE=85000468; PubMed=6433976;
RA Pohl G., Kaelstroom M., Bergsdorf N., Wallen P., Joernvall H.;
RT "Tissue plasminogen activator: peptide analyses confirm an indirectly
RT derived amino acid sequence, identify the active site serine residue,
RT establish glycosylation sites, and localize variant differences.";
RL Biochemistry 23:3701-3707(1984).
RN [11]
RP SEQUENCE OF 33-52 AND 311-330.
RC TISSUE-Melanoma;
RX MEDLINE=83209620; PubMed=6682760;
RA Wallen P., Pohl G., Bergsdorf N., Raanby M., Ny T., Joernvall H.;
RT "Purification and characterization of a melanoma cell plasminogen
RT activator";
RL Eur. J. Biochem. 132:681-686(1983).
RN [12]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE-Umbilical vein;
RX MEDLINE=90192129; PubMed=2107528;
RA Siebert P.D., Fong K.;
RT "Variant tissue-type plasminogen activator (PLAT) cDNA obtained from
RT human endothelial cells";
RL Nucleic Acids Res. 18:1086-1086(1990).
RN [13]
RP STRUCTURE OF CARBOHYDRATES.
RX MEDLINE=90092112; PubMed=2513186;
RA Pfeiffer G., Schmidt M., Strube K.-H., Geyer R.;
RT "Carbohydrate structure of recombinant human uterine tissue
RT plasminogen activator expressed in mouse epithelial cells.";
RL Eur. J. Biochem. 186:273-286(1989).
RN [14]
RP CARBOHYDRATE-LINKAGE SITE THR-96.
RX MEDLINE=91159408; PubMed=1900431;
RA Harris R.J., Leonard C.K., Guzzetta A.W., Spellman M.W.;
RT "Tissue plasminogen activator has an O-linked fucose attached to
RT threonine-61 in the epidermal growth factor domain.";
RL Biochemistry 30:2311-2314(1991).
RN [15]
RP DISULFIDE BONDS IN KRINGLE 2.
RX MEDLINE=91244765; PubMed=1645336;
RA Vlahos C.J., Wilhelm O.G., Hassell T., Jaskunas S.R., Bang N.U.;
RT "Disulfide pairing of the recombinant kringle-2 domain of tissue
RT plasminogen activator produced in Escherichia coli.";
RL J. Biol. Chem. 266:10070-10072(1991).
RN [16]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF CATALYTIC DOMAIN.
RX MEDLINE=96200985; PubMed=8613982;
RA Lamba D., Bauer M., Huber R., Fischer S., Rudolph R., Kohnert U.,
RA Bode W.;
RT "The 2.3 A crystal structure of the catalytic domain of recombinant
RT two-chain human tissue-type plasminogen activator.";
RL J. Mol. Biol. 258:117-135(1996).
RN [17]
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF CATALYTIC DOMAIN.
RX MEDLINE=97449126; PubMed=9305622;
RA Renatus M., Engh R.A., Stubbs M.T., Huber R., Fischer S., Kohnert U.,
RA Bode W.;
RT "Lysine 156 promotes the anomalous proenzyme activity of tPA: X-ray
RT crystal structure of single-chain human tPA.";
RL EMBO J. 16:4797-4805(1997).
RN [18]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF KRINGLE 2.
RX MEDLINE=92118803; PubMed=1310033;
RA de Vos A., Oltch M.H., Kelley R.F., Padmanabhan K., Tulinsky A.,
RA Westbrook M.L., Kossiakof A.A.;
RT "Crystal structure of the kringle 2 domain of tissue plasminogen
RT activator at 2.4-A resolution.";
RL Biochemistry 31:270-279(1992).
RN [19]
RP STRUCTURE BY NMR OF KRINGLE 2.
RX MEDLINE=90122799; PubMed=2558718;
RA Byeon I.-J.L., Kelley R.F., Llinas M.;
RT "1H NMR structural characterization of a recombinant kringle 2 domain
RT from human tissue-type plasminogen activator.";
RL Biochemistry 28:9350-9360(1989).
RN [20]
RP STRUCTURE BY NMR OF KRINGLE 2.
RX MEDLINE=91200042; PubMed=1901789;
RA Byeon I.-J.L., Kelley R.F., Llinas M.;
RT "Kringle-2 domain of the tissue-type plasminogen activator. 1H-NMR
RT assignments and secondary structure.";
RL Eur. J. Biochem. 197:155-165(1991).
RN [21]
RP STRUCTURE BY NMR OF KRINGLE 2.
RX MEDLINE=92106329; PubMed=1762144;
RA Byeon I.-J.L., Llinas M.;
RT "Solution structure of the tissue-type plasminogen activator kringle
RT 2 domain complexed to 6-aminohexanoic acid an antifibrinolytic
RT drug";
RL J. Mol. Biol. 222:1035-1051(1991).
RN [22]
RP STRUCTURE BY NMR OF 38-85.
RX MEDLINE=92292163; PubMed=1602484;

Downing A.K., Driscoll P.C., Harvey T.S., Dudgeon T.J., Smith B.O.,
 Baron M., Campbell I.D., "Solution structure of the fibrin binding finger domain of
 tissue-type plasminogen activator determined by 1H nuclear magnetic
 resonance.",
 J Mol. Biol. 225:821-833(1992).
 RA STRUCTURE BY NMR OF 36-126.
 RX MEDLINE=96027104; PubMed=7582899;
 RA Smith B.O., Downing A.K., Driscoll P.C., Dudgeon T.J., Campbell I.D.,
 RA "The solution structure and backbone dynamics of the fibronectin type
 I and epidermal growth factor-like pair of modules of tissue-type
 plasminogen activator.",
 J Mol. Biol. 225:821-833(1992).
 CC -!- FUNCTION CONVERTS THE ABUNDANT, BUT INACTIVE, ZMOGEN PLASMINOGEN
 TO PLASMIN BY HYDROLYZING A SINGLE ARG-VAL BOND IN PLASMINOGEN. BY
 CONTROLLING PLASMIN-MEDIATED PROTEOLYSIS, IT PLAYS AN IMPORTANT
 ROLE IN TISSUE REMODELING AND DEGRADATION, IN CELL MIGRATION AND
 MANY OTHER PHYSIOLOGICAL EVENTS.
 CC -!- CATALYTIC ACTIVITY: SPECIFIC CLEAVAGE OF ARG-|-VAL BOND IN
 CC ZMOGENOGEN.
 CC -!- SUBUNIT: BINDS TO FIBRIN WITH HIGH AFFINITY. THIS INTERACTION

Query Match 43.2%; Score 38; DB 1; Length 562;
 Best Local Similarity 44.4%; Pred. No. 74;
 Matches 8; Conservative 5; Mismatches 3; Indels 2; Gaps 1;
 QY 1 DWDTSALA--PYLQWE 16
 DB 150 NWNSSALAKPYSGRRPD 167

RESULT 26
 UGST_ORYSA STANDARD; PRT; 609 AA.
 AC Q42968;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GRANULE-BOUND GLYCAGEN [STARCH] SYNTHASE PRECURSOR (EC 2.4.1.11).
 GN WAXY, glaberrima (African rice).
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae;
 CC Oryza.
 OX NCBI_TaxID=4538;
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. JAPONICA, AND CV. HANFENG;
 RX MEDLINE=92134825; PubMed=1685658;
 RA Umesa M., Ohtsubo H., Ohtsubo E.,
 RA "Diversification of the rice waxy gene by insertion of mobile DNA
 elements into introns.",
 Jpn. J. Genet. 66:569-586(1991).
 CC -!- CATALYTIC ACTIVITY: FOR THE SYNTHESIS OF AMYLOSE IN ENDOSPERM.
 CC -!- FUNCTION: REQUIRED FOR THE SYNTHESIS OF AMYLOSE + (1,4-BETA-D-GLUCOSYL)[N] - UDP +
 CC (1,4-BETA-D-GLUCOSYL)[N+1].
 CC -!- PATHWAY: STARCH BIOSYNTHESIS.
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST, GRANULE-BOUND.
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCAGEN SYNTHASE
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC EMBL: D10472; BAA01272.1;
 CC Glycogen biosynthesis; Transferrase; Glycosyltransferase;
 KW Transit peptide; Chloroplast; Starch biosynthesis.

TRANSIT 1 77 CHLOROPLAST (BY SIMILARITY).
 CHAIN 78 609 GRANULE-BOUND GLYCAGEN [STARCH] SYNTHASE.
 BINDING 97 97 UDP-GLUCOSE (BY SIMILARITY).
 SQ SEQUENCE 609 AA; 66475 MW; C228BFB9C407FA5 CRC64;
 Query Match 43.2%; Score 38; DB 1; Length 609;
 Best Local Similarity 63.6%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 DWDTSALADYL 11
 DB 234 DWHGTPLASYL 244

RESULT 27
 UGST_ORYSA STANDARD; PRT; 609 AA.
 AC P19395; Q43013;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GRANULE-BOUND GLYCAGEN [STARCH] SYNTHASE PRECURSOR (EC 2.4.1.11).
 GN WAXY OR WX.
 CC Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae;
 CC Oryza.
 OX NCBI_TaxID=4530;
 RN SEQUENCE FROM N.A.
 RP STRAIN=Seed;
 RX MEDLINE=92322986; PubMed=1377969;
 RA Ogasawara S.,
 RA "Nucleotide sequence of a long cDNA from the rice waxy gene.",
 Plant Mol. Biol. 19:513-516(1992).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. JAPONICA TAICHUNG 65; TISSUE=Seedling;
 RA Hirano H.Y., Sano Y.,
 RA "Molecular characterization of the waxy locus of rice (Oryza
 sativa L.)",
 Plant Cell Physiol. 32:989-997(1991).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. JAPONICA, AND CV. HANFENG;
 RX MEDLINE=91016948; PubMed=2216792;
 RA Wang M.M., Wu Z.L., Xing F.G., Zheng F.G., Guo X.L., Zhang W.G.,
 RA "Nucleotide sequence of rice waxy gene.",
 Nucleic Acids Res. 18:5898-5898(1990).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. INDICA;
 RX Wang M.M., Hong M.M.,
 RX "Nucleotide sequence of a rice waxy gene.",
 Submitted (MAY 1993), to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE OF 153-343 FROM N.A.
 RX MEDLINE=91200672; PubMed=2016064;
 RA Shinada H., Tada Y.,
 RA "Rapid isolation of a rice waxy sequence: a simple PCR method for the
 analysis of recombinant plasmids from intact Escherichia coli
 cells.",
 Gene 98:243-248(1991).
 CC -!- FUNCTION: REQUIRED FOR THE SYNTHESIS OF AMYLOSE IN ENDOSPERM.
 CC -!- CATALYTIC ACTIVITY: UDP-GLUCOSE + (1,4-ALPHA-D-GLUCOSYL)[N] - UDP
 CC + (1,4-ALPHA-D-GLUCOSYL)[N+1].
 CC -!- PATHWAY: STARCH BIOSYNTHESIS.
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST, GRANULE-BOUND.
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCAGEN SYNTHASE
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X62134; CAA44065.1; -
DR EMBL; X58228; CAA41186.1; -
DR EMBL; X53694; CAA37732.1; -
DR EMBL; X65183; CAA46294.1; -
DR EMBL; M55039; AAA33918.1; -
DR PIR; JQ0703; JQ0703.
DR PIR; S22519; S22519.
DR PIR; S30485; S30485.
DR PIR; JQ2224; JQ2224.

KW Glycogen biosynthesis; Transferase; Glycosyltransferase;
transit peptide; Chloroplast; Starch biosynthesis.

FT TRANSIT 1 77
FT CHAIN 78 609 GRANULE-BOUND GLYCAGEN [STARCH] SYNTHASE.
FT BINDING 97 97 UDP-GLUCOSE (BY SIMILARITY).
FT CONFLICT 247 247 N -> T (IN REF. 5).
FT CONFLICT 250 250 P -> T (IN REF. 5).
FT CONFLICT 415 415 P -> S (IN REF. 4).
SQ SEQUENCE 609 AA; 66476 MW; C225DBF6F12072C5 CRC64;

Query Match 43.2%; Score 38; DB 1; Length 609;
Best Local Similarity 63.6%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 DWDTSALAPYL 11
DB 234 DWHGTGLASYL 244

RESULT 28
AMVH_SACDI

ID AMVH_SACDI STANDARD; PRT; 767 AA.

AC P04065; Q92314;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUCOAMYLASE S1 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-GLUCOSIDASE)
DE (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE) (GAL).
GN STAI OR DEX2 OR MAL5.
OS Saccharomyces diastaticus (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=41870;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 60709;
RX MEDLINE=85104778; PubMed=3918017;
RA Yamashita I., Suzuki K., Fukui S.;
RT "Nucleotide sequence of the extracellular glucoamylase gene STAI in
the yeast Saccharomyces diastaticus.";
RN J. Bacteriol. 161:567-573(1985).
RN [2]
RP SEQUENCE OF 1-54 FROM N.A.
RA Yamashita I., Suzuki K., Sakuzo F.;
RT "Proteolytic processing of glucoamylase in the yeast Saccharomycetes
cerevisiae.";
RN Agric. Biol. Chem. 50:475-482(1986).
RN [3]
RP SEQUENCE OF 1-64 FROM N.A.
RA Shima H., Inui M., Akada R., Yamashita I.;
RT "Upstream regions of the yeast glucoamylase gene which are required
for efficient transcription.";
RN Agric. Biol. Chem. 53:749-755(1989).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-
GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS
WITH RELEASE OF BETA-D-GLUCOSE.

-1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X02649; CAA36487.1; ALT_INIT.

DR EMBL; D00428; BAA00332.1; -

DR PIR; A21896; ALBYG.

DR HSP; P08017; LAYX.

DR InterPro; IPR000165; -

DR Pfam; PF00723; Glyco_hydro_15; 1.

DR PROSITE; PS00820; GLUCOAMYLASE; 1.

KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;

KW Signal; Multigene family.

FT SIGNAL 1 21

FT CHAIN 22 767 GLUCOAMYLASE S1.

FT DOMAIN 22 347 SER/THR-RICH.

FT DOMAIN 348 691 H SUBUNIT.

FT DOMAIN 692 767 Y SUBUNIT.

FT BINDING 455 455 SUBSTRATE (BY SIMILARITY).

FT ACT_SITE 518 518 CATALYTIC BASE (BY SIMILARITY).

FT ACT_SITE 521 521 GENERAL ACID CATALYST (BY SIMILARITY).

FT ACT_SITE 522 522 INTERACT WITH SUBSTRATES (BY SIMILARITY).

FT CARBOHYD 35 35 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 308 308 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 322 322 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 434 434 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 513 513 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 546 546 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 650 650 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 720 720 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 741 741 N-LINKED (GLCNAC. .) (POTENTIAL).

SQ SEQUENCE 767 AA; 82488 MW; A5F29E2427EDB593 CRC64;

Query Match 43.2%; Score 38; DB 1; Length 767;

Best Local Similarity 46.7%; Pred. No. 1e-02;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 WDTLSALAPYLGTQEE 16

DB 333 WDSSALEEWLQKK 347

RESULT 29

PCGB_BOVIN

ID PCGB_BOVIN STANDARD; PRT; 912 AA.

AC Q28062;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE BREVICAN CORE PROTEIN PRECURSOR.

GN BCAN.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Brain;

RX MEDLINE=94193597; PubMed=8144512;

RA Yamada H., Watanabe K., Shimonaka M., Yamaguchi Y.;

RT "Molecular cloning of brevicin, a novel brain proteoglycan of the

aggrecan/versican family.";


```
SQ SEQUENCE 2292 AA; 255495 MW; 8540D0EB1437EBD4 CRC64;

Query Match 43.2%; Score 38; DB 1; Length 2292;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWDTSALAPY 10
   ||::: | ||
Db 1959 DWESATLIPY 1968

RESULT 31
POLG EMCVD STANDARD; PRT; 2292 AA.
AC P17594;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)].
OS Encephalomyocarditis virus (strain emc-d diabetogenic).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Cardiovirus.
OX NCBI_TaxID=12106;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89243189; PubMed=2541543;
RA Bae Y.S., Eun H.M., Yoon J.W.;
RT "Genomic differences between the diabetogenic and nondiabetogenic
RT variants of encephalomyocarditis virus.";
RL Virology 170:282-287(1989).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M22458; AAA43034.1; -.
CC PIR; A31473; GNMYED.
CC HSSP; P12296; IMEC.
CC MEROPS; C03.009; -.
CC DR MEROPS; U29.001; -.
CC DR InterPro; IPR000605; -.
CC DR InterPro; IPR001205; -.
CC DR InterPro; IPR001676; -.
CC Pfam; PF00680; RNA_dep_RNA_pol; 1.
CC Pfam; PF00910; RNA_helicase; 1.
CC DR Pfam; PF00073; rhv; 3.
CC DR Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Helicase; Thiol protease; Myristate.
FT PROPEP 1 67
FT CHAIN 68 137 COAT PROTEIN VP4 (RHQ).
FT CHAIN 138 393 COAT PROTEIN VP2 (BETA).
FT CHAIN 394 624 COAT PROTEIN VP3 (GAMMA).
FT CHAIN 625 901 COAT PROTEIN VP1 (ALPHA).
FT CHAIN 902 1058 CORE PROTEIN P2A (G).
FT CHAIN 1059 1194 CORE PROTEIN P2B (I).
FT CHAIN 1195 1519 CORE PROTEIN P2C (F).
FT CHAIN 1520 1607 CORE PROTEIN P3A.
FT CHAIN 1608 1627 GENOME-LINKED PROTEIN VPG (H).

SQ SEQUENCE 2292 AA; 255426 MW; F2B0627B0F444107 CRC64;

Query Match 43.2%; Score 38; DB 1; Length 2292;
Best Local Similarity 50.0%; Pred. No. 2.9e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWDTSALAPY 10
   ||::: | ||
Db 1959 DWESATLIPY 1968

RESULT 32
MTX2 CAEEL STANDARD; PRT; 260 AA.
AC P34599;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE METAXIN 2 HOMOLOG.
GN ZC97.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL OUTER MEMBRANE (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE METAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L14714; AAA28180.1; -.
CC PIR; S44885; S44885.
CC WormPep; ZC97.1; CE00360.
KW Hypothetical protein; Mitochondrion; Outer membrane; Transport;
KW Protein transport.
SQ SEQUENCE 260 AA; 29752 MW; 4A9E46E937C3CFEE CRC64;

Query Match 42.6%; Score 37.5; DB 1; Length 260;
Best Local Similarity 53.3%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1 DW-DTSALAPYLGTQ 14
   || | | | | | | |
```

```

DB 22 DWEDSVLTFPLNDQ 36
RESULT 33
YNIF_AZOB
AC NITRILE-118 AA.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HYPOTHETICAL 12.3 KDA PROTEIN IN NIFK-NIFY INTERGENIC REGION.
OS Azospirillum brasilense.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC Azospirillum.
OC NCBI_TaxID=192;
RN (1) NCBI_TaxID=192;
RP SEQUENCE FROM N.A.
RX MEDLINE=92370074; PubMed=1823284;
RA Passaglia L.M.P., Nunes C.P., Zaha A., Schrank I.S.;
RT "The nifHDK operon in the free-living nitrogen-fixing bacterium
RT Azospirillum brasilense sequentially comprises genes H, D, K, an 353
RT Braz J Med Biol Res. 24:649-675(1991).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC EMBL; M64344; AAB02345.1;
CC PIR; S27476; S27476.
CC SWISSPROT; S27476; S27476.
CC NCBI_TaxID=118;
CC SEQUENCE 118 AA; 12323 MW; 54783E0F3F4D8D53 CRC64;
Query Match 42.0%; Score 37; DB 1; Length 118;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 DWDTSAAP 9
DB 107 DWDTSAAP 115
RESULT 34
ID NHA_PSECL STANDARD; PRT; 199 AA.
AC P27764;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NITRILE HYDRATASE SUBUNIT ALPHA (EC 4.2.1.84) (NITRILASE) (NHASE).
OS Pseudomonas chlororaphis (Pseudomonas fluorescens biotype D).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OC NCBI_TaxID=333;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-25 AND 163-177.
RX STRAIN=821193202; PubMed=2013568;
RA Nishiyama M., Horinouchi S., Kobayashi M., Nagasawa T., Yamada H.,
RA Beppu T.;
RT "Cloning and characterization of genes responsible for metabolism of
RT nitrile compounds from Pseudomonas chlororaphis B23."
RL J. Bacteriol. 173:2465-2472(1991).
CC -1- FUNCTION: NHASE CATALYZES THE HYDRATION OF VARIOUS NITRILE
CC COMPOUNDS. NHASE CATALYZES THE HYDRATION OF NITRILE COMPOUNDS TO
CC ACRYLAMIDE. IS NOW BEING DEVELOPED USING SOME OF THESE ENZYMES.
CC -1- CATALYTIC ACTIVITY: AN ALIPHATIC AMIDE -> A NITRILE + H(2)O.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC EMBL; U02341; AAA91964.1;
CC HSSP; P17676; P17676.
CC PIR; P17676; P17676.
CC NCBI_TaxID=118;
CC SEQUENCE 199 AA; 21986 MW; P380F6C428C251EE CRC64;
Query Match 42.0%; Score 37; DB 1; Length 199;
Best Local Similarity 70.0%; Pred. No. 39;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 2 WDTSAAPYL 11
DB 155 WDTSAAPYL 164
RESULT 35
ID NHA_PSECL STANDARD; PRT; 206 AA.
AC P33651;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SUPEROXIDE DISMUTASE [MN] (EC 1.15.1.1).
OS Nocardia asteroides.
OC Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Nocardia.
OC NCBI_TaxID=1824;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96060854; PubMed=7590304;
RA Alcendor D.J., Chapman G.D., Beaman B.L.;
RT "Isolation, sequencing and expression of the superoxide dismutase-
RT encoding gene (sod) of Nocardia asteroides strain GUH-2."
RL Gene 164:143-147(1995).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELL. CATALYZES THE DISMUTATION OF SUPEROXIDE RADICALS TO O(2) + H(2)O(2).
CC CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) -> O(2) + H(2)O(2).
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC EMBL; U02341; AAA91964.1;
CC HSSP; P17676; P17676.
CC PIR; P17676; P17676.
CC NCBI_TaxID=118;
CC SEQUENCE 199 AA; 21986 MW; P380F6C428C251EE CRC64;
Query Match 42.0%; Score 37; DB 1; Length 199;
Best Local Similarity 70.0%; Pred. No. 39;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 2 WDTSAAPYL 11
DB 155 WDTSAAPYL 164

```

```

Best Local Similarity 72.7%; Pred. No. 42;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 TSALAPYLGTQ 14
DB 17 TSLLAHLGTQ 27

RESULT 37
YP97_CAEEL STANDARD; PRT; 283 AA.
AC Q09245;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL 33.3 KDA PROTEIN C28H8.7 IN CHROMOSOME III.
GN C28H8.7
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Miller N., Waterston R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U20861; AAA62295.1; -.
DR WormPep; C28H8.7; CE01829.
KW Hypothetical protein.
SQ SEQUENCE 283 AA; 33295 MW; CD8941ABE66CE9F8 CRC64;

Query Match 42.0%; Score 37; DB 1; Length 283;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 WDTSAAPYLGTQ 15
DB 159 WDYGSLANYLNQSE 172

RESULT 38
THTR_RAT
ID THTR_RAT STANDARD; PRT; 295 AA.
AC P24329;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE THIOSULFATE SULFURTRANSFERASE (EC 2.8.1.1) (RHODANESE) (FRAGMENT).
GN TST.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RC MEDLINE=91207296; PubMed=2018478;
RA Weiland K.L., Dooley T.P.;
RT "Molecular cloning, sequencing and characterization of cDNA to rat
RT liver rhodanese, a thiosulphate sulphurtransferase.";
RL Biochem. J. 275:227-231(1991).
RN [2]
RP MUTAGENESIS.

```


DR PDB; 1RHS; 21-JAN-98.
DR PDB; 2ORA; 01-AUG-96.
DR PDB; 10RB; 15-OCT-95.
DR PDB; 1BOH; 27-APR-99.
DR PDB; 1BOI; 27-APR-99.
DR InterPro; IPR001307; -.
DR InterPro; IPR001763; -.
DR Pfam; PF00581; Rhodanese; 2.
DR PROSITE; PS00380; RHODANESE_1; 1.
DR PROSITE; PS00683; RHODANESE_2; 1.
KW Transferase; Mitochondrion; 3D-structure.
FT INIT_MET 0 0
FT DOMAIN 1 142 A DOMAIN.
FT DOMAIN 143 158 HINGE.
FT DOMAIN 159 296 B DOMAIN.
FT ACT_SITE 186 186 MAY PLAY A ROLE IN SUBSTRATE BINDING.
FT ACT_SITE 247 247
FT ACT_SITE 248 248
FT ACT_SITE 249 249
FT ACT_SITE 249 249
FT VARIANT 1 2
FT MUTAGEN 186 186
FT MUTAGEN 249 249
FT CONFLICT 99 99
FT CONFLICT 214 214
FT CONFLICT 219 219
FT STRAND 9 10
FT HELIX 12 21
FT TURN 22 22
FT STRAND 24 24
FT TURN 25 27
FT STRAND 28 32
FT TURN 38 40
FT HELIX 43 47
FT TURN 48 49
FT STRAND 51 51
FT TURN 53 54
FT STRAND 56 57
FT TURN 60 61
FT TURN 66 67
FT HELIX 77 87
FT TURN 91 92
FT STRAND 94 97
FT TURN 102 103
FT HELIX 108 117
FT TURN 118 118
FT STRAND 123 126
FT TURN 127 128
FT HELIX 129 136
FT TURN 137 137
FT STRAND 141 141
FT TURN 158 159
FT STRAND 161 162
FT TURN 164 165
FT HELIX 166 173
FT STRAND 177 180
FT HELIX 184 188
FT TURN 189 189
FT STRAND 203 203
FT TURN 205 206
FT STRAND 208 209
FT TURN 212 215
FT STRAND 216 216
FT TURN 218 219
FT STRAND 222 222
FT HELIX 225 234
FT TURN 235 236
FT STRAND 239 240
FT STRAND 243 246
FT TURN 252 253
FT HELIX 254 263

FT TURN 264 264
FT TURN 266 267
FT STRAND 269 271
FT TURN 272 274
FT HELIX 275 281
FT HELIX 284 286
FT STRAND 288 288
SQ SEQUENCE 296 AA; 33164 MW; C8769696FA6AC111 CRC64;

Query Match 42.0%; Score 37; DB 1; Length 296;
Best Local Similarity 57.1%; Pred. NO. 58;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 16
Db 180 DSRAGRYLGTQPE 193

RESULT 40
THTR_MOUSE
ID THTR_MOUSE STANDARD; PRT; 296 AA.
AC P52196;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE THIOSULFATE SULFURTRANSFERASE (EC 2.8.1.1) (RHODANESE).
GN TST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=96074596; PubMed=7488186;
RA Dooley T.P., Nair S.K., Garcia R.E., Courtney B.C.;
FT "Mouse rhodanese gene (Tst): cDNA cloning, sequencing, and
recombinant protein expression.";
RL Biochem. Biophys. Res. Commun. 216:1101-1109(1995).
CC -!- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES AND CYANIDE
DETOXIFICATION.
CC -!- CATALYTIC ACTIVITY: THIOSULFATE + CYANIDE = SULFITE + THIOCYANATE.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -!- TISSUE SPECIFICITY: FOUND IN NUMEROUS TISSUES.
CC -!- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR
CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,
THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.
CC -!- SIMILARITY: BELONGS TO THE RHODANESE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; U35741; AAC52342.1; -.
DR HSSP; P00586; IRHD.
DR SWISS-2DPAGE; P52196; MOUSE.
DR MGD; MGI:98852; Tst.
DR InterPro; IPR001307; -.
DR InterPro; IPR001763; -.
DR Pfam; PF00581; Rhodanese; 2.
DR PROSITE; PS00380; RHODANESE_1; 1.
DR PROSITE; PS00683; RHODANESE_2; 1.
KW Transferase; Mitochondrion.
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 1 142 A DOMAIN.
FT DOMAIN 143 158 HINGE.
FT DOMAIN 159 296 B DOMAIN.

FT ACT_SITE 186 186 MAY PLAY A ROLE IN SUBSTRATE BINDING (BY
FT ACT_SITE 247 247 BY SIMILARITY)
FT ACT_SITE 248 248 SUBSTRATE (THIOSULFATE) BINDING
FT ACT_SITE 249 249 (BY SIMILARITY)
FT SEQUENCE 296 AA; 33334 MW; 82069DD080F9A555A CRC64;
Query Match 42.0%; Score 37; DB 1; Length 296;
Best Local Similarity 57.1%; Pred. No. 58;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
OY 3 DWSALAPYLGTQEE 16
DB 180 DSRAGRYLGTQEE 193
RESULT 41
ID HCV3_HSV11 STANDARD; PRT; 302 AA.
AC 000115;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE HYPOTHETICAL GENE 3 PROTEIN.
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Unclassified Herpesviridae.
OX NCBI_TaxID=10401;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92087430; PubMed=1727613;
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus";
RL Virology 186:9-14(1992).
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
DR EMBL; M75136; AAA88184.1;
DR EMBL; M75136; AAA88106.1;
DR PIR; D36786; D36786.
KW Hypothetical protein.
SQ SEQUENCE 302 AA; 32282 MW; B03FE6B3180F19FD CRC64;
Query Match 42.0%; Score 37; DB 1; Length 302;
Best Local Similarity 42.9%; Pred. No. 59;
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
OY 1 DWDTSALAPYLGTQ 14
DB 84 DWDSPASFPVTGAE 97
RESULT 42
ID HPRK_TREPA STANDARD; PRT; 319 AA.
AC 083600; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE HPR(SER) KINASE/PHOSPHATASE (EC 2.7.1.-) (EC 3.1.3.-).
OS HPRK OR PTSK OR TP0591.
OC HPRK OR PTSK OR TP0591.

OS Treponema pallidum.
OX Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OY NCBI_TaxID=160;
DB [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Childs R.A., Adams J.T.,
RA McDonald L., Attachan C., Bowman C., Cote M.D., Fujii C., Garland S.,
RA Venter J.C., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA "Complete genome sequence of Treponema pallidum, the syphilis
RT Spirochete";
RT Science 281:375-388(1998).
CC -1- FUNCTION: THIS KINASE REGULATES CARBOHYDRATE UPTAKE
CC AND METABOLISM. IT IS INVOLVED IN PHOSPHORYLATING/DEPHOSPHORYLATING SER-46 OF THE
CC PHOSPHORYL CARRIER PROTEIN (HPR) OF THE PHOSPHOENOLPYRUVATE-
CC DEPENDENT SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS) IN RESPONSE TO
CC CYTOSOLIC METABOLITE LEVELS AS A DEVICE FOR THE MODULATION OF
CC CARBON CATABOLITE REPRESSION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PTSK FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
DR EMBL; AE001234; AAC65566.1;
DR TIGR; TP0591;
DR KW Multifunctional enzyme; Transferase; Serine/threonine-protein kinase;
FT ATP-binding; Hydrolase.
FT NP_BIND 152 133 ATP (POTENTIAL)
SQ SEQUENCE 319 AA; 35548 MW; 5ADF40485574B12D CRC64;
Query Match 42.0%; Score 37; DB 1; Length 319;
Best Local Similarity 46.7%; Pred. No. 62;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
OY 1 DWDTSALAPYLGTQEE 15
DB 234 EWNSSKAYDLRLGTQEE 248
RESULT 43
ID PKX1_HUMAN STANDARD; PRT; 358 AA.
AC P51817;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE PROTEIN KINASE PKX1 (EC 2.7.1.-).
GN PRKX OR PKXAS (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis, Brain;
RA Klink A., Schiebel K., Winkelman M., Rao E., Horsthemke B.,
RA Lueddecke H.-J., Clausen U., Scherer G., Rappold G.;
RT "The human protein kinase gene PKX1 on Xp22.3 displays xp/yp homology
RT and is a site of chromosomal instability";
RL Hum. Mol. Genet. 4:869-878(1995).
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN ADULT AND FETAL BRAIN, KIDNEY

•

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:55:32 ; Search time 72.61 Seconds
(without alignments)
29.154 Million cell updates/sec

Title: US-09-439-313-562

Perfect score: 88

Sequence: 1 DWDTSAAPVLTQEE 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database :

SPTREMBL_16:

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organella:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.rodent:*
- 12: sp.unclassified:*
- 13: sp.vertebrate:*
- 14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	46	52.3	491	2 P94289	P94289 bacillus ci
2	45	51.1	685	2 Q9K1Q0	Q9K1Q0 neisseria m
3	45	51.1	685	2 Q9JWP0	Q9JWP0 neisseria m
4	45	51.1	831	2 O50076	O50076 clostridium
5	45	51.1	847	2 Q9I4U2	Q9I4U2 pseudomonas
6	43	48.9	207	2 Q9F9R1	Q9F9R1 mycobacteri
7	43	48.9	275	2 Q9I567	Q9I567 pseudomonas
8	43	48.9	402	2 Q9S118	Q9S118 synechocyst
9	43	48.9	459	2 Q9X7Q9	Q9X7Q9 streptomyce
10	42	47.7	227	3 P79022	P79022 candida sp.
11	42	47.7	490	2 Q9PLR4	Q9PLR4 chlamydia m
12	42	47.7	563	14 Q9N2J9	Q9N2J9 herv-h/enh6
13	41.5	47.2	1044	2 Q9F4D6	Q9F4D6 bifidobacte
14	41	46.6	217	1 O58748	O58748 pyrococcus
15	41	46.6	332	2 Q9S3X3	Q9S3X3 streptomyce
16	41	46.6	343	2 O85977	O85977 sphingomona
17	41	46.6	343	2 Q923U6	Q923U6 sphingomona
18	41	46.6	359	6 Q28450	Q28450 toxodonta a
19	41	46.6	410	6 Q28556	Q28556 orycteropus

RESULT 1

P94289 ID P94289 PRELIMINARY; PRT; 491 AA.

AC P94289;
DT 01-MAY-1997 (TREMUREL. 03, Created)
DT 01-MAY-1997 (TREMUREL. 03, Last sequence update)
DT 01-MAR-2001 (TREMUREL. 16, Last annotation update)

DE CHITINASE C PRECURSOR.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.

NCBI_TaxID=1397;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-WL-12;

RA Alam M.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-WL-12;

RA Alam M., Nikaldou N., Tanaka H., Watanabe T.;

RT *cloning and sequencing of chlc gene of Bacillus circulans WL-12 and

ALIGNMENTS

Q9Y9T8 aeropyrum p
Q9N1T0 ornithohyn
Q9L4X3 streptomyce
O16266 caenorhabdi
Q911R6 pseudomonas
Q9M6J3 aquilegia s
Q9LH14 arabisdopsi
Q59326 clostridium
Q9F111 arabidopsi
Q9X645 arabidopsi
Q9SDA6 arabidopsi
Q9VSC9 drosophila
P72316 rhodospiril
Q15019 homo sapien
Q9HSG5 halobacteri
Q53813 mycobacteri
Q9LGN9 oryza sativ
Q53803 mycobacteri
P78203 escherichia
P79827 oncorhynchu
Q9L0K0 streptomyce
Q9GJT2 sus scrofa
O54397 rhodococcus
O16739 caenorhabdi
Q58813 pyrococcus
Q9HP38 halobacteri
Q55481 synechocyst
O46520 sus scrofa
Q9LH39 arabidopsi
Q29585 archaeoglob
Q9I134 pseudomonas
Q9HYC7 pseudomonas
Q9PFV8 xylella fas
Q9KT69 vibrio chol
Q9HZE4 pseudomonas
Q94928 homo sapien
Q83422 mengo encep
Q9FDJ9 bacteroides
Q97218 leishmania
Q9SDN1 prunus dulc
Q83368 murine leuk
Q9GPR2 dictyosteli
Q9HJ08 thermoplasm
O65523 arabidopsi
P73809 synechocyst
O22451 zea mays (m

```

RT relationship of its product to some other chitinases and chitinase-
RL like chitinases.
CC J. Ferment. Bioceng. 80:454-461(1995).
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC -1- HYDROLASES).
DR EMBL: D89568; BAAL3974.1; -.
DR HSSP: P07254; ICTN; -.
DR InterPro: IPR001573; -.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR PROSITE: PS01095; CHITINASE_18; 1.
DR Glycosidase; Hydrolase; Signal.
FT SIGNAL 1
FT CHAIN 42
FT POTENTIAL 41
FT SIGNAL 41
FT CHAIN 42
FT POTENTIAL 41
SQ SEQUENCE 491 AA; 53447 MW; 31A98122AF9BCBEC CRC64;

Query Match 52.3%; Score 46; DB 2; Length 491;
Best Local Similarity 80.0%; Pred. No. 9.7;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 WDTSALAPYL 11
DB 318 WDSSAQAPYL 327

RESULT 2
OY Q9K100 PRELIMINARY; PRT; 685 AA.
AC Q9K100: 2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE METHIONYL-TRNA SYNTHETASE.
GN NM000300
NCBI_TaxID=313
CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RT "Complete genome sequence of a serogroup A strain of Neisseria meningitidis (serogroup B)."
SQ SEQUENCE 287:1809-1815(2000).

Query Match 51.1%; Score 45; DB 2; Length 685;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

OY 1 DWDTSAAPYL 12
DB 237 DWDISRDAPYFG 248

RESULT 3
OY Q9JWPO PRELIMINARY; PRT; 685 AA.
AC Q9JWPO: 2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10).
GN METG OR NMA0275.
NCBI_TaxID=63699;
CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
RT "Complete genome sequence of a serogroup A strain of Neisseria meningitidis (serogroup A)."
SQ SEQUENCE 22491.1; 2000.

Query Match 51.1%; Score 45; DB 2; Length 685;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 DWDTSAAPYL 12
DB 237 DWDISRDAPYFG 248

RESULT 4
OY O50076 PRELIMINARY; PRT; 831 AA.
AC O50076: 2000 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMBlrel. 06, Last annotation update)
DE CHITINASE B (EC 3.2.1.14) (CHITINASE)
GN CHIB
OS Clostridium parapatricum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
OX NCBI_TaxID=29363;
RN [1] J. Ferment. Bioceng. 80:454-461(1995).
RX STRAIN=M21.
CC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
RT "Cloning, sequencing, and expression of the gene encoding Clostridium parapatricum chitinase ChIB and analysis of the functions of novel chitinase-like 1397206-7314(1997)
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
EMBL: AB001874; BAA23796.1; -.
HSSP: P07254; ICTN.

```

RC STRAIN-K-10;
RA Liu X., Feng Z., Cirillo J., Barletta R.G.;
RT "Mycobacterium paratuberculosis manganese superoxide dismutase";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF180816; AAG09425.1; -
SQ SEQUENCE 207 AA; 23030 MW; EDA8C2EB40ED428D CRC64;

Query Match 48.9%; Score 43; DB 2; Length 207;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0

Qy 1 DWDTSALAPYLGTQ 14
||| :||| :||| :
Db 10 DWDYAALPEPHISGQ 23

RESULT 7
Q91567 ID Q91567 PRELIMINARY; PRT; 275 AA.
AC Q91567;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE DE HYPOTHETICAL PROTEIN PA0878.
GN PA0878.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004522; AAG04267.1; -
KW Hypothetical protein.
SQ SEQUENCE 275 AA; 30521 MW; 5DC3223738335DFE CRC64;

Query Match 48.9%; Score 43; DB 2; Length 275;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0

Qy 5 SALAPYLGTQEE 16
||| :||| :||| :
Db 4 SAFAPWIGRQEE 15

RESULT 8
Q55118 ID Q55118 PRELIMINARY; PRT; 402 AA.
AC Q55118;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE DE HYPOTHETICAL 43.9 KDA PROTEIN.
GN SLI0408.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA Tabata S.;
RP SEQUENCE FROM N.A.

DR InterPro: IPR001223; -
DR InterPro: IPR003610; -
DR Pfam: PF00704; Glyco_Hydro_18; 2.
DR SMART: SM00495; ChtBD3; 1.
DR KW Hydrolase; Glycosidase.
SQ SEQUENCE 831 AA; 90021 MW; 6982E0B45B72AD31 CRC64;

Query Match 51.1%; Score 45; DB 2; Length 831;
Best Local Similarity 64.3%; Pred. No. 26;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0

Qy 2 WDTSAALPYLGTQEE 15
||| :||| :||| :
Db 400 WDDSAKAPLYLNPE 413

RESULT 5
Q91402 ID Q91402 PRELIMINARY; PRT; 847 AA.
AC Q91402;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE DE PROBABLE PENICILLIN AMIDASE.
GN PA1032.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong K.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004535; AAG04421.1; -
DR InterPro: IPR002692; -
DR Pfam: PF01804; Penicill.amidase; 4.
SQ SEQUENCE 847 AA; 94069 MW; 141B2C463E3071C6 CRC64;

Query Match 51.1%; Score 45; DB 2; Length 847;
Best Local Similarity 52.9%; Pred. No. 26;
Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

Qy 1 DW--DTSALAPYLGTQEE 15
||| :||| :||| :
Db 721 EWQSDSKMAPYLGA 737

RESULT 6
Q9F9R1 ID Q9F9R1 PRELIMINARY; PRT; 207 AA.
AC Q9F9R1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE DE SUPEROXIDE DISMUTASE.
GN SOD.
OS Mycobacterium paratuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RP SEQUENCE FROM N.A.


```

DE DNA GYRASE, SUBUNIT A.
GN TC0031.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
  White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
  Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
  RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
  RA Eisen J., Fraser C.M.;
  "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
  RT pneumoniae AR39."; 28:1397-1406(2000).
RL Nucleic Acids Res. 28:1397-1406(2000).
DR EMBL; AF002271; AAF38922.1; -;
DR TIGR; TC0031; -;
DR InterPro; IPR002205; -;
DR Pfam; PF00521; DNA_topoisom; 2;
SQ SEQUENCE 490 AA; 54999 MW; 28BDB36FC2541B00 CRC64;

Query Match 47.7%; Score 42; DB 2; Length 490;
Best Local Similarity 80.0%; Pred. No. 47;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 TSALAPYLGT 13
DB 378 TDALAPFLGT 387
  | | | | |
  | | | | |

RESULT 12
Q9N2J9 ID Q9N2J9 PRELIMINARY; PRT; 563 AA.
AC Q9N2J9
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE ENVELOPE PROTEIN.
GN ENV.
OS HERV-H/env60.
OC Viruses; Retroviridae.
OX NCBI_TaxID=129526;
RN [1]
RP SEQUENCE FROM N.A.
RA de Parseval N., Casella J.F., Gressin L., Heidmann T.;
  "Characterization of the three envelopes of the HERV-H family
  RT encompassing the ISU domain and evolutionary history in primates.";
  Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ289710; CAB94193.2; -;
DR InterPro; IPR002050; -;
DR Pfam; PF00429; ENV_polyprotein; 2;
SQ SEQUENCE 563 AA; 61884 MW; 961E3E2FEDB5C8D4 CRC64;

Query Match 47.7%; Score 42; DB 14; Length 563;
Best Local Similarity 53.8%; Pred. No. 55;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGT 13
DB 84 DWGTSPVSPHLRT 96
  | | | | |
  | | | | |

RESULT 13
Q9F4D6 ID Q9F4D6 PRELIMINARY; PRT; 1044 AA.
AC Q9F4D6
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)

```

```

DE BETA-GALACTOSIDASE (EC 3.2.1.23).
OS Bifidobacterium bifidum.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OX Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=1681;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 20215;
RA Poulsen U., Stougaard P.;
  Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ224434; CAC14565.1; -;
DR EMBL; AJ224434; CAC14565.1; -;
KW Hydrolase; Glycosidase.
SQ SEQUENCE 1044 AA; 117182 MW; 81C2475AD4C72178 CRC64;

Query Match 47.2%; Score 41.5; DB 2; Length 1044;
Best Local Similarity 52.9%; Pred. No. 1.3e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

QY 2 WDTSA---LAPYLGTQE 15
DB 935 WDATAKSMAPYLWVQE 951
  | | | | |
  | | | | |

RESULT 14
O58748 ID O58748 PRELIMINARY; PRT; 217 AA.
AC O58748
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE 217AA LONG HYPOTHETICAL AROM PROTEIN.
GN PH1049;
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
  Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
  Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
  RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kushida N., Oguchi A.,
  RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
  RA Masuchi Y., Shizuya H., Kikuchi H.;
  "Complete sequence and gene organization of the genome of a hyper-
  RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
  DNA Res. 5:55-76(1998).
RL EMBL; AF000004; BAA30147.1; -;
DR InterPro; IPR002034; -;
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
SQ SEQUENCE 217 AA; 24303 MW; E23C491C0D57E87B CRC64;

Query Match 46.8%; Score 41; DB 1; Length 217;
Best Local Similarity 54.5%; Pred. No. 28;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 6 ALAPYLGTQEE 16
DB 154 SVSPYVGTTEE 164
  | | | | |
  | | | | |

RESULT 15
Q9S3X3 ID Q9S3X3 PRELIMINARY; PRT; 332 AA.
AC Q9S3X3
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE SPCN.
GN SPCN.

```

OS Streptomyces spectabilis.
 OC Bacteriales; Streptomycetaceae; Actinobacteridae;
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
 RN NCBI_TaxID=68270;
 RP [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=NRRL 2494;
 RA "Jennsen D., Volf J.N.;
 RT Streptomyces spectabilis";
 RT Streptomyces spectabilis";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF170704; AAD50455.1;
 SQ SEQUENCE 332 AA; 36128 MW; 76703C681FF9A39A CRC64;

Query Match 46.6%; Score 41; DB 2; Length 332;
 Best Local Similarity 77.8%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 DWTSALAP 9
 DB 243 DWDIVGLAP 251

RESULT 16

ID O85977 PRELIMINARY; PRT; 343 AA.
 AC O85977; 1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE 4-HYDROXY-2-OXOVALERATE ALDOLASE.
 GN XYLK.
 OS Sphingomonas aromaticivorans.
 OG Plasmid pNLI.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
 OC Actinomycetales; Streptomycetaceae; Streptomyces.
 RN NCBI_TaxID=48935;
 RP [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=F199;
 RA Romine M.F., Stillwell L.C., Wong K.-K., Thurston S.J., Sisk E.C.,
 RA Jensen C.W., Gaasterland F., Saffer J.D., Fredrickson R.K.;
 RT Sphingomonas aromaticivorans strain F199; a catabolic plasmid from Sphingomonas
 RT aromaticivorans.
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF079317; AAD03993.1;
 DR InterPro: IPR000891;
 DR Pfam: PF00682; HMGL-like; 1.
 SQ SEQUENCE 343 AA; 36859 MW; 465022794B8C9C9 CRC64;

Query Match 46.6%; Score 41; DB 2; Length 343;
 Best Local Similarity 33.3%; Pred. No. 47;
 Matches 10; Conservative 2; Mismatches 4; Indels 14; Gaps 1;

QY 1 DWD-----TSALAPVLGTQEE 16
 DB 70 DWDWTEAAADVKNVAVLTLLVPGIGTAAE 99

RESULT 17

ID O92306 PRELIMINARY; PRT; 343 AA.
 AC O92306;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DE 4-HYDROXY-2-OXOVALERATE ALDOLASE (4-HYDROXY-2-KETOVALERATE ALDOLASE).
 GN Sphingomonas chungbukensis.
 OS Sphingomonas chungbukensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;

OC Sphingomonas
 RN NCBI_TaxID=56193;
 RP [1]
 RC SEQUENCE FROM N.A.
 RA STRAIN=DJ77;
 RA MEDLINE=99182313; Pubmed=10080921;
 RA Hwang S.-I., Kim S.-J., Kim C.-K., Kim Y.-C., Kim Y.-C.;
 RT 4-Hydroxy-2-oxovalerate aldolase from Sphingomonas sp. DJ77 and their
 RT evolutionary implications";
 RL Biochem. Biophys. Res. Commun. 256:469-473(1999).
 DR EMBL: AF079317; AAD03993.1;
 DR InterPro: IPR000891;
 DR Pfam: PF00682; HMGL-like; 1.
 SQ SEQUENCE 343 AA; 36863 MW; A6A02A7493D53825 CRC64;

Query Match 46.6%; Score 41; DB 2; Length 343;
 Best Local Similarity 33.3%; Pred. No. 47;
 Matches 10; Conservative 2; Mismatches 4; Indels 14; Gaps 1;

QY 1 DWD-----TSALAPVLGTQEE 16
 DB 70 DWDWTEAAADVKNVAVLTLLVPGIGTAAE 99

RESULT 18

ID O28450 PRELIMINARY; PRT; 359 AA.
 AC Q28450;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 GN IRBP-RETINOID BINDING PROTEIN (FRAGMENT).
 OS Loxodonta africana (African elephant).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Proboscidea; Elephantidae; Loxodonta.
 RN NCBI_TaxID=9785;
 RP [1]
 RC SEQUENCE FROM N.A.
 RA MEDLINE=96304325; Pubmed=8660440;
 RA Stanhope M.J., Smith M.R., Waddell V.G., Porter C.A., Shivji M.S.,
 RA Goodman M.;
 RT "Mammalian evolution and the interphotoreceptor retinoid binding
 RT protein, (IRBP) gene: convincing evidence for several superordinal
 RT clades." Evol. 43:83-92(1996).
 DR EMBL: U48711; AAB18658.1;
 DR InterPro: IPR003581;
 DR SMART: SM00245; TSPC; 1.
 FT NON_TER 359
 FT NON_TER 359
 SQ SEQUENCE 359 AA; 39216 MW; 2C759A4621C0D312 CRC64;

Query Match 46.6%; Score 41; DB 6; Length 359;
 Best Local Similarity 46.7%; Pred. No. 50;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 WPTSALAPVLGTQEE 16
 DB 201 WEGSGVLPVGTPEE 215


```
RESULT 19
Q28556 PRELIMINARY; PRT; 410 AA.
ID Q28556
AC Q28556
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE INTERPHOTORECEPTOR RETINOID BINDING PROTEIN (FRAGMENT).
GN IRBP.
OS Oryctolopus afer (Aardvark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Tubulidentata; Orycteropodidae; Orycteropus.
OX NCBI_TaxID=9818;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96304325; PubMed=8660440;
RA Stanhope M.J., Smith M.R., Waddell V.G., Porter C.A., Shiomi M.S.,
RA Goodman M.;
RT "Mammalian evolution and the interphotoreceptor retinoid binding
RT protein (IRBP) gene: convincing evidence for several superordinal
RT clades.";
RL J. Mol. Evol. 43:83-92(1996).
DR EMBL; U48712; AAB18659.1; -
DR InterPro; IPR003581; -
DT SMART; SM00245; TSPC; 1.
FT NON_TER 1 1
FT NON_TER 410 410
SQ SEQUENCE 410 AA; 44964 MW; 791C6D9F1CB31FAA CRC64;

Query Match 46.6%; Score 41; DB 6; Length 410;
Best Local Similarity 46.7%; Pred. No. 58;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16
   | : | : | : | : |
Db 252 WEGSGVLPCVGTPEE 266

RESULT 20
Q9Y9T8 PRELIMINARY; PRT; 427 AA.
ID Q9Y9T8
AC Q9Y9T8
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE 427AA LONG HYPOTHETICAL L-ASPARAGINASE.
GN APE2200.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999)
DR EMBL; AF000063; BAA81212.1; -
DR InterPro; IPR000267; -
DR Pfam; PF00710; Asparaginase; 1.
DR PRINTS; PR00139; ASNGINASE.
DR PROSITE; PS00917; ASN_GLN_ASE_2; 1.
SQ SEQUENCE 427 AA; 46373 MW; B932AD12884ED923 CRC64;

Query Match 46.6%; Score 41; DB 6; Length 427;
Best Local Similarity 46.7%; Pred. No. 4.3e+02;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16
   | : | : | : | : |
Db 905 WDTEACPVATTKDE 919

RESULT 22
Q9L4X3 PRELIMINARY; PRT; 9477 AA.
ID Q9L4X3
AC Q9L4X3
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE NYSI.
DE NYSI.
GN NYSI.
OS Streptomyces noursei.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1971;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
```

RA Brattaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
 RA Valia S., Zolchev S.B., Glycine antifungal antibiotic nystatin in
 RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
 RT deduction of the biosynthetic pathway.";
 RL Chem. Biol. 7:395-403(2000).
 CC -1- SIMILARITY: TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR)
 CC FAMILY: 012; AAF71766.1; --
 CC EMBL: AF26612; AAF71766.1; --
 DR IPR000255; --
 DR InterPro: IPR000794; --
 DR InterPro: IPR001005; --
 DR InterPro: IPR001227; --
 DR InterPro: IPR001899; --
 DR InterPro: IPR002198; --
 DR IPR002198; KdsShort; 3;
 DR Pfam: PF00109; KdsShort; synt; 6;
 DR Pfam: PF00109; KdsShort; synt; 6;
 DR Pfam: PF00550; pp-binding; 6;
 DR Pfam: PF00698; Acyl_transf; 6;
 DR PROSITE: PS00075; ACP_DOMAIN; 6;
 DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 6;
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1;
 DR PROSITE: PS00312; GRAM_POS_ANCHORING; UNKNOWN_1;
 DR PROSITE: PS00012; PHOSPHOPANTHETINE; 6;
 DR Oxidoreductase; Phosphopantetheine; Transferase;
 KW OXIDOREDUCTASE; PHOSPHOPANTHETINE; TRANSFERASE;
 SQ SEQUENCE 9477 AA; 988151 MW; 0D2BCA5D7B265483 CRC64;

 Query Match 45.5%; Score 41; DB 2; Length 9477;
 Best Local Similarity 56.2%; Pred. No. 1;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

 QY 1 DWDTSALAPYLGTQEE 16
 Db 4267 DGDVSAALTAALGTDED 4282

 RESULT 23
 ID 016266 PRELIMINARY; PRT; 258 AA.
 AC O16266; 1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 RT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE F40A3.6 PROTEIN.
 GN F40A3.6
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 CC [1]_TaxID=6239;
 RN [1] DWDTSALAPYLGTQEE 16
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Grant S., Hawkins P., Hillier N., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaughan M., Vaughan K., Waterston R.,
 RA Watson A., Welinkstock L., Wilkinson-Sproat C., Wodman P.,
 RT elegans." Contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-BRISTOL N2;
 RA Geisel C., Bradshaw H., Keppler D.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF016423; AAB65325.1; -- 7D6579D97B2E1611 CRC64;
 SQ SEQUENCE 258 AA; 30243 MW;

 Query Match 45.5%; Score 40; DB 5; Length 258;
 Best Local Similarity 63.6%; Pred. No. 51;
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

 QY 1 DWDTSALAPYL 11
 Db 201 DFDSSPLSPIL 211

 RESULT 24
 ID 0911R6 PRELIMINARY; PRT; 294 AA.
 AC 0911R6; 2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 RT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL PROTEIN PA2201.
 GN PA2201.
 OS Pseudomonas aeruginosa.
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC [1]_TaxID=287;
 RN [1] DWDTSALAPYL 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20437337; PubMed=10984043;
 RA Hoyer C.A., Brinkman C.F., Erwin R.J., Mizoguchi S.D., Warren P.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.I., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wory Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen."
 RT Nature 406:959-964(2000).
 DR EMBL: AE004646; AAG05589.1; --
 KW Hypothetical protein.
 SQ SEQUENCE 294 AA; 32850 MW; EA6F273152A14996 CRC64;

 Query Match 45.5%; Score 40; DB 2; Length 294;
 Best Local Similarity 66.7%; Pred. No. 59;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

 QY 3 DTSALAPYLGTQ 14
 Db 218 DAQALAPYLSQ 229

 RESULT 25
 ID 09M6J3 PRELIMINARY; PRT; 392 AA.
 AC 09M6J3; 2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 RT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE PHYTOCHROME C (FRAGMENT).
 GN PHYC.
 OS Aquilegia sp. SM-1999.
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; Ranunculales; Ranunculaceae; Aquilegia.
 CC [1]_TaxID=112038;
 RN [1] DTSALAPYLGTQ 14
 RP SEQUENCE FROM N.A.
 RA Mathews S., Donoghue M.J.;

RT "Root of the angiosperm phylogenetic tree inferred from duplicated
 RT phytochrome genes.";
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF190067; AAF26313.1; -;
 DR 01-MAR-2001 (TREMBlrel. 15, Last annotation update)
 DR InterPro: IPR001294; -;
 DR InterPro: IPR003018; -;
 DR Pfam: PF00360; phytochrome; 1.
 DR PRINTS: PR01033; PHYTOCHROME.
 DR PROSITE: PS00245; PHYTOCHROME_1; 1.
 DR PROSITE: PS00446; PHYTOCHROME_2; 1.
 DR SMART: SM00065; GAF; 1.
 FT NON_TER 1
 FT NON_TER 392 392
 SQ SEQUENCE 392 AA; 43847 MW; EBF4A728C1EEB529 CRC64;

Query Match 45.5%; Score 40; DB 10; Length 392;
 Best Local Similarity 58.3%; Pred. No. 82;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 DWDTSALAPYL 12
 DB 75 EWKISDLEPYL 86

RESULT 26
 Q9LH14 PRELIMINARY; PRT; 440 AA.
 AC Q9LH14;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE SERINE CARBOXYPEPTIDASE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX PubMed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL: AF002047; BAB03131.1; -;
 DR InterPro: IPR001563; -;
 DR Pfam: PF00450; serine_carpept; 2.
 DR PRINTS: PR00724; CRBOXYPTASEC.
 KW Carboxypeptidase.
 SQ SEQUENCE 440 AA; 49890 MW; D70C6F5E26ED8C3C CRC64;

Query Match 45.5%; Score 40; DB 10; Length 440;
 Best Local Similarity 53.8%; Pred. No. 93;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQE 15
 DB 354 DHDMLTPYVGTQD 366

RESULT 27
 Q59326 PRELIMINARY; PRT; 482 AA.
 ID Q59326

AC Q59326;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE CHITINASE PRECURSOR (EC 3.2.1.14) (CHITODEXTRINASE) (1,4-BETA-POLY-N-
 DE ACETYLGLUCOSAMINIDASE) (POLY-BETA-GLUCOSAMINIDASE).
 GN CHIA.
 OS Clostridium thermocellum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 1237;
 RA Fuchs K.P., Schwarz W.H., Staudenbauer W.L.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF N-
 CC ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
 CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
 CC HYDROLASES).

DR EMBL: Z68924; CAA93150.1; -;
 DR HSSP: P07254; ICTN.
 DR InterPro: IPR001223; -;
 DR InterPro: IPR001579; -;
 DR InterPro: IPR002048; -;
 DR InterPro: IPR002105; -;
 DR Pfam: PF00404; Dockerin_1; 2.
 DR Pfam: PF00704; Glyco_Hydro_18; 2.
 DR PROSITE: PS01095; CHITINASE_18; 1.
 DR PROSITE: PS00448; CLOS_CELLULOSE_RPT; 2.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN_2.
 KW Glycosidase; Hydrolase; Signal.
 FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 482
 SQ SEQUENCE 482 AA; 55028 MW; BF047D8DE3588DF5 CRC64;

Query Match 45.5%; Score 40; DB 2; Length 482;
 Best Local Similarity 50.0%; Pred. No. 1e+02;
 Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 WDTSAALAPYLGTQE 15
 DB 341 WDEYAMVPYLWNP 354

RESULT 28
 Q9FFL1 PRELIMINARY; PRT; 490 AA.
 ID Q9FFL1
 AC Q9FFL1;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE SIMILARITY TO CATIONIC AMINO ACID TRANSPORTER 1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=97471969; PubMed=9330910;
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT pl clones.";
 RL DNA Res. 4:215-230(1997).
 DR EMBL: AB005237; BAB09657.1; -;
 SQ SEQUENCE 490 AA; 53693 MW; E478743784CFE309 CRC64;

```

Query Match          45.5%; Score 40; DB 10; Length 490;
Best Local Similarity 50.0%; Pred. NO. 1.1e+02;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

      2 WDTSLALPYLGTOE 15
      | | | | | | |
      120 WVTLMGPIYWGFOO 133

RESULT 29
-----
Q9XE45; PRELIMINARY; PRT; 732 AA.
01-NOV-1999 (TREMBLrel. 12, Created)
01-NOV-1999 (TREMBLrel. 12, Last sequence update)
01-MAR-2001 (TREMBLrel. 16, Last annotation update)
TSM2.3 HAVE REVERSE TRANSCRIPTASE.
Arabidopsis thaliana (Mouse-ear cross).
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,
Carrster A.J., Cresay T.H., Buell C.R., Town C.D., Nierman W.C.,
Fraser C.M., Venter J.C.;
"Arabidopsis thaliana chromosome II BAC T5M2 genomic sequence.";
Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
1- TRANSCRIPTASE) RNA-DIRECTED DNA POLYMERASE (REVERSE
EMBL: AC007730; RAD37021.1; -
Interpro: IPR000477; -
Pfam: PF00078; rvt; 2.
RNA-directed DNA polymerase.
SEQUENCE 732 AA; 83450 MW; 3D58BF00D56E8F2E CRC64;

Query Match          45.5%; Score 40; DB 10; Length 732;
Best Local Similarity 41.7%; Pred. NO. 1.7e+02;
Matches .5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

      2 WDTSLALPYLCTG 13
      | | | | | | |
      500 WDMATIAPYISS 511

RESULT 30
-----
Q9SDA6 PRELIMINARY; PRT; 957 AA.
01-MAY-2000 (TREMBLrel. 13, Created)
01-MAY-2000 (TREMBLrel. 16, Last sequence update)
01-MAY-2000 (TREMBLrel. 16, Last annotation update)
PUTATIVE DISEASE RESISTANCE PROTEIN.
AV2G17050.
Arabidopsis thaliana (Mouse-ear cross).
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
NCBI_TaxId=3702;
SEQUENCE FROM N.A.
STRAIN=CV. COLUMBIA;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Hallonborg G., Jorgensen
```

```
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;
RL Science 287:2185-2195(2000).
DR EMBL; AE003556; RAF50494.1; -.
DR FlyBase; FBgn0035833; CG7565.
DR InterPro; IPR000561; -.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR SMART; SM00181; EGF; 1.
SQ SEQUENCE 1069 AA; 117310 MW; 2BA50A62EEB463DB CRC64;

Query Match 45.5%; Score 40; DB 5; Length 1069;
Best Local Similarity 46.7%; Pred. No. 2.5e+02;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGTQEE 16
Db 741 WDLAVVVKYLWTRDE 755

RESULT 32
P72316 PRELIMINARY; PRT; 1265 AA.
ID P72316;
DC MEDLINE-97047978; PubMed-8892819;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CARBON MONOXIDE-INDUCED HYDROGENASE OPERON COOM PROTEIN.
GN COOM.
OS Rhodospirillum rubrum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC Rhodospirillum.
OX NCBI_TaxID=1085;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=URL;
RA MEDLINE-97047978; PubMed-8892819;
RA Fox D.J., He Y., Shelper D., Roberts G.P., Ludden P.W.;
RT *Characterization of the region encoding the CO-induced hydrogenase of
RT Rhodospirillum rubrum.*;
RL J. Bacteriol. 178:6200-6208(1996).
CC -!- FUNCTION: MAY BE THE MAJOR MEMBRANE ANCHOR FOR THE ENERGY
CC -!- CONSERVING FOR CARBON MONOXIDE-INDUCED HYDROGENASE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
DR EMBL; U65510; AAC45116.1; -.
DR InterPro; IPR001750; -.
DR Pfam; PF00361; oxidored_q1; 5.
KW Transmembrane.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 227 247 POTENTIAL.
FT TRANSMEM 278 298 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 474 494 POTENTIAL.
FT TRANSMEM 518 538 POTENTIAL.
FT TRANSMEM 580 600 POTENTIAL.
FT TRANSMEM 603 623 POTENTIAL.
FT TRANSMEM 704 724 POTENTIAL.
FT TRANSMEM 728 748 POTENTIAL.
FT TRANSMEM 756 776 POTENTIAL.
FT TRANSMEM 784 804 POTENTIAL.
FT TRANSMEM 842 862 POTENTIAL.
FT TRANSMEM 877 897 POTENTIAL.
FT TRANSMEM 962 982 POTENTIAL.
FT TRANSMEM 1001 1021 POTENTIAL.
FT TRANSMEM 1093 1113 POTENTIAL.
FT TRANSMEM 1143 1163 POTENTIAL.
```

```
FT TRANSMEM 1187 1207 POTENTIAL.
SQ SEQUENCE 1265 AA; 132997 MW; D681B9299A9E2C5E CRC64;

Query Match 45.5%; Score 40; DB 2; Length 1265;
Best Local Similarity 54.5%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WDTSALAPYLGT 12
Db 229 WDAFVLLPYVG 239

RESULT 33
O15019 PRELIMINARY; PRT; 2047 AA.
ID O15019;
AC O15019;
DC MEDLINE-97349984; PubMed-9205841;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE KIAA0301 (DJ12208.4) (KIAA0301) (FRAGMENT).
GN KIAA0301 OR DJ12208.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE-97349984; PubMed-9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT *Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.*;
RL DNA Res. 4:141-150(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB002299; BAA20761.1; -.
DR EMBL; AL096678; CAB86660.1; -.
DR InterPro; IPR002035; -.
FT NON_TER 1
SQ SEQUENCE 2047 AA; 231843 MW; AA1EDD0452CECB4E CRC64;

Query Match 45.5%; Score 40; DB 4; Length 2047;
Best Local Similarity 43.8%; Pred. No. 5.2e+02;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 DWDTSALAPYLGTQEE 16
Db 1308 DYDENEVDVPYHNGQEK 1323

RESULT 34
Q9HSG5 PRELIMINARY; PRT; 575 AA.
ID Q9HSG5;
AC Q9HSG5;
DC MEDLINE-97047978; PubMed-8892819;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE VNG0243C.
GN VNG0243C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20504483; PubMed-11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
```

RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Madsen D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Eberhard H., Lowe T.M., Liang P., Riley M., Nodwell T., Dassarma S.,
RA Genome sequence of *Halobacterium salinarum* R1.
RA EMBL: AF004988; AAG18841.1; ...
RA InterPro: IPR001656; ...
RA Pfam: PF01142; UPF0024; 1.
RA SEQUENCE 575 AA; 63009 MW; C0172D520F687808 CRC64;
SQ
Query Match 44.3%; Score 39.5; DB 1; Length 575;
Best Local Similarity 50.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 3; Mismatches 4; Indels 1; Gaps 1;
QY 1 DWDTSAALAPYLQEE 16
DB 338 DWEEAAMA-YLGAPTE 352
RESULT 35
ID O53813 PRELIMINARY; PRT; 81 AA.
AC O53813
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE HYPOTHEICAL 8.7 KDA PROTEIN.
GN RV0750 OR MV041.24.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Badcock K., Basham D., Baskin D., Berrington J., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Seeger K., Skellton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence."
RA EMBL: AL021958; CAA17517.1; ...
RA Tuberculin; Rv0750; ...
RA Hypothetical protein.
RA SEQUENCE 81 AA; 8719 MW; A02ED63B68DAFEA CRC64;
SQ
Query Match 44.3%; Score 39; DB 2; Length 81;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 DWDTSAALAPYL 11
DB 60 WYTOGLAPYL 69
RESULT 36
ID O9LGN9 PRELIMINARY; PRT; 84 AA.
AC O9LGN9
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE P0684C01.6 PROTEIN.
GN Oryza sativa (Pico).
OS Oryza sativa (Pico).
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RC P0684C01.6; ...
RT P0684C01.6; ...
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002487; BAB07945.1; ...
DR EMBL: AP002487; BAB07945.1; ...
SQ SEQUENCE 84 AA; 9085 MW; 33CA61DF0A6CCE0C CRC64;
Query Match 44.3%; Score 39; DB 10; Length 84;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 DWDTSAALAP 9
DB 62 DWDTSAALAP 70
RESULT 37
ID O53803 PRELIMINARY; PRT; 175 AA.
AC O53803
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE HYPOTHEICAL 19.5 KDA PROTEIN.
GN RV0740 OR MV041.14.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Badcock K., Basham D., Baskin D., Berrington J., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Seeger K., Skellton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrett B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence."
RA EMBL: AL021958; CAA17517.1; ...
RA Tuberculin; Rv0750; ...
RA Hypothetical protein.
RA SEQUENCE 175 AA; 19452 MW; CD8EB84AE02A6723 CRC64;
SQ
Query Match 44.3%; Score 39; DB 2; Length 175;
Best Local Similarity 70.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 DWDTSAALAPYL 11
DB 154 WYTOGLAPYL 163
RESULT 38
P78203

ID P78203 PRELIMINARY; PRT; 179 AA.
AC P78203;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE BENZENE 1,2-DIOXYGENASE ALPHA SUBUNIT (EC 1.14.12.3) (FRAGMENT).
GN BEDC1.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampo G., Satoh Y., Sivasundaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT *Construction of a contiguous 874-kb sequence of the Escherichia coli
-K12 genome corresponding to 50.0-68.8 min on the linkage map and
analysis of its sequence features.*;
RL DNA Res. 4:91-113(1997).
DR EMBL: D90883; BAA16433.1; -.
DR HSSP: P23094; 1NDO.
DR InterPro: IPR001281; -.
DR InterPro: IPR001663; -.
DR Pfam: PF00355; Rieske; 1.
DR PRINTS: P00090; RINGDIOLGNASE.
DR PROSITE: PS00570; RING-HYDROXYL_ALPHA; 1.
FT NON_TER 179 179
SQ SEQUENCE 179 AA; 20454 MW; 78FDA1562601573D CRC64;

Query Match 44.3%; Score 39; DB 2; Length 179;
Best Local Similarity 64.3%; Pred. No. 51;
Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 1 DWDTSAPGLRDYLG 12
:||||| | |||
DB 155 NWDTSAPGLRDYLG 168

RESULT 39
P79827
ID P79827 PRELIMINARY; PRT; 196 AA.
AC P79827;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CATHEPSIN L (EC 3.4.22.15) (FRAGMENT).
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Nickel X.F., An H., Li F., Seymour T.A., Morrissey M.T., Barnes D.W.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: SPECIFICITY CLOSE TO THAT OF PAPAIN.
DR EMBL: U61296; AAB37252.1; -.
DR HSSP: P07711; 1CJL.
DR MEROPS: C01.032; -.
DR InterPro: IPR000169; -.
DR InterPro: IPR000668; -.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS: PR00705; PAPAIN.
DR PROSITE: PS00640; THIOLESTERASE_ASN; 1.
DR PROSITE: PS00639; THIOLESTERASE_HIS; 1.

KW Hydrolase; Thiol protease.
FT NON_TER 1
SQ SEQUENCE 196 AA; 21537 MW; 2AB15321E1E4779E CRC64;

Query Match 44.3%; Score 39; DB 13; Length 196;
Best Local Similarity 50.0%; Pred. No. 56;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 DTSALAPYLGTQEE 16
|||:|:|:|:
DB 58 DTEESYPVGTDED 71

RESULT 40
Q910K0
ID Q910K0 PRELIMINARY; PRT; 209 AA.
AC Q910K0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PUTATIVE SECRETED PROTEIN.
GN SCD40A.lilc.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT *A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.*;
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL161691; CAB81856.1; -.
SQ SEQUENCE 209 AA; 21951 MW; 63BB4EEFB9718BC CRC64;

Query Match 44.3%; Score 39; DB 2; Length 209;
Best Local Similarity 66.7%; Pred. No. 61;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WDTSAAPY 10
|:|:|:|:
DB 121 WSTAAPY 129

RESULT 41
Q9GJT2
ID Q9GJT2 PRELIMINARY; PRT; 282 AA.
AC Q9GJT2;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ESTERASE D.
GN PIGESD OR ESD.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-ONCHICHIDA S., Kajii E.;
RL "Sus scrofa cDNA for Esterase D (phenotype A), complete cds."
RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RA Tsuchida S., Omi T., Ikemoto S.;
RC Cloning of a cDNA.
RT Cloning of a cDNA.
DR EMBL; AB038362; BAB11922.1; -
DR EMBL; AF254785; AAG17630.1; -
SQ SEQUENCE 282 AA; 31482 MW; 6284689291591FBA CRC64;

Query Match 44.3%; Score 39; DB 6; Length 282;
Best Local Similarity 50.0%; Pred. No. 85;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
QY 2 WDTSALAPYLGTQEE 15
DB 183 WGGKAFSGYLGTDE 196

RESULT 42
ID O54397 PRELIMINARY; PRT; 326 AA.
AC O54397;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-1998 (TREMBLrel. 07, Last annotation update)
DE F45C12.8 PROTEIN.
OS Rhodococcus erythropolis.
OG Plasmid pBD2.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.
OX NCBI_TaxID=1833;
RN [1] SEQUENCE FROM N.A.
RC Kessler M., Averhoff B., Gottschalk G.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U83846; AAB94318.1; -
KW Plasmid.
SQ SEQUENCE 326 AA; 34702 MW; 9CD534D5B2A19298 CRC64;

Query Match 44.3%; Score 39; DB 2; Length 326;
Best Local Similarity 77.8%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;
QY 4 TSALAPYLIG 12
DB 190 TSLAPYLIG 198

RESULT 43
ID O16739 PRELIMINARY; PRT; 337 AA.
AC O16739; 1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE F45C12.8 PROTEIN.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae;
OC Eukaryota; Ecdysozoa; Ecdyzerinac; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC STRAIN-BRISTOL N2;
SQ SEQUENCE FROM N.A.
MEDLINE=94150718; Pubmed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copest T., Cooper J.,
RA Craxton A., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins J., Hilder N., Lattelle P.,
RA Jones M., Kershaw J., Kilmurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thiermy-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinstein L., Wilkinson-Sproat J., Woldman P., III of C.
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RL Johnson D., Kramer J., Keppler D.;
RN Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RC STRAIN-BRISTOL N2;
RL Waterston R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016684; AAB66207.1; -
DR InterPro: IPR001810; -
DR InterPro: IPR002900; -
DR Pfam: PF01827; DUF38; 1;
DR Pfam: PF01827; DUF38; 2;
DR PROSITE: PS00181; FBOX; 1;
DR SMART: SM00256; FBOX; 1;
SQ SEQUENCE 337 AA; 39534 MW; FB45E9DEAB390DB4 CRC64;

Query Match 44.3%; Score 39; DB 5; Length 337;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 1 DWTSALAPYLGTQEE 16
DB 286 DMNTSLKSVFGTQDE 301

RESULT 44
ID O58813 PRELIMINARY; PRT; 346 AA.
AC O58813;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE 34624 LONG HYPOTHETICAL ASPARTOKINASE.
GN PH1086.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC KAWARABAYASI Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
EX MEDLINE=98344137; Pubmed=9679194;
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya A., Ohnuki Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki K., Kusuda N., Oguchi A.,
RA Aoki K.-I., Toshiyawa T., Kikuchi H.,
RA "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RT Thermophilic archaeobacterium, Pyrococcus horikoshii OT3.*;
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000004; BAA30185.1; -
DR InterPro: IPR001048; -
DR InterPro: IPR001341; -
DR PROSITE: PS00124; ASPARTOKINASE; 1;
DR PROSITE: PS00124; ASPARTOKINASE; UNKNOWN.1;
SQ SEQUENCE 346 AA; 38543 MW; F5958F0ECFA7470D CRC64;

Query Match 44.3%; Score 39; DB 1; Length 346;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 DTSALAPYL 11
|||:|:|
DB 80 DTSSLSPLYL 88

RESULT 45

Q9HP38 PRELIMINARY; PRT; 347 AA.
AC Q9HP38;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE ALCOHOL DEHYDROGENASE.
GN ADH4 OR VNG1821G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Ballig N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlshroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AF005083; AAG20032.1; -;
DR InterPro; IPR002085; -;
DR Pfam; PF00107; adh_zinc; 1.
SQ SEQUENCE 347 AA; 36719 MW; 5352453D854E26E1 CRC64;

Query Match 44.3%; Score 39; DB 1; Length 347;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DWTSALAP 9
|||:|:|
DB 140 DWETAAAP 148

Search completed: June 28, 2001, 11:55:36
Job time: 157 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:54:12 ; Search time 66.55 Seconds
(without alignments)
10.931 Million cell updates/sec

Title: US-09-439-313-558
Perfect score: 68
Sequence: 1 AGLCLPDRPLE 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : A_Geneseq_0601.*

- 1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	100.0	553	19	AAW71869
2	68	100.0	553	19	AAW69385
3	68	100.0	553	21	AAW28527
4	68	100.0	553	21	AAW82002
5	45	66.2	191	21	AAW42089
6	41	60.3	15	21	AAW98874
7	41	60.3	15	21	AAW98970
8	41	60.3	59	21	AAW21202
9	41	60.3	266	21	AAW21201
10	41	60.3	291	19	AAW63682
11	41	60.3	293	21	AAW58875

RESULT 1

AAW71869 AAW71869 standard; Protein; 553 AA.

XX	AC	AAW71869;
XX	DT	06-JAN-1999 (first entry)
XX	DE	Amino acid encoded by prostate tumour clone I1-12.
XX	KW	Prostate; cancer; tumour; vaccine; immunogen; clone.
XX	OS	Homo sapiens.
XX	XX	

ALIGNMENTS

12	41	60.3	568	16	AAW65496	Marek's disease vi
13	41	60.3	712	21	AAW21204	Human HER-2/neu fu
14	41	60.3	729	22	AAU00020	Human plexin prote
15	41	60.3	919	21	AAW21203	Human HER-2/neu fu
16	41	60.3	1200	21	AAW21208	Human HER-2/neu pr
17	41	60.3	1255	17	AAW01111	HER-2/neu protein.
18	41	60.3	1255	20	AAW92406	Human HER-2/neu pr
19	41	60.3	1255	21	AAW21198	Human HER-2/neu pr
20	41	60.3	1255	21	AAW84780	Amino acid sequenc
21	41	60.3	1255	21	AAW92620	Human heregulin 2
22	41	60.3	1255	22	AAW60167	HER2 transgene pla
23	41	60.3	1433	14	AAW39568	Sequence of c-erbB
24	41	60.3	2135	22	AAU00019	Human plexin prote
25	40	58.8	122	12	AAW15105	HCG/DLH chimera, D
26	40	58.8	328	17	AAW05504	HCW Toledo strain
27	40	58.8	485	21	AAW31001	Arabidopsis thalia
28	40	58.8	486	21	AAW31000	Arabidopsis thalia
29	40	58.8	522	20	AAW38566	Neisseria meningit
30	40	58.8	522	20	AAW38568	Neisseria meningit
31	40	58.8	523	20	AAW38567	Neisseria meningit
32	40	58.8	526	21	AAW74891	Neisseria meningit
33	40	58.8	533	21	AAW74890	Neisseria gonorrhoe
34	40	58.8	533	21	AAW74892	Neisseria meningit
35	40	58.8	538	21	AAW30999	Arabidopsis thalia
36	39	57.4	161	21	AAW22586	zea mays protein f
37	39	57.4	172	21	AAW33891	Arabidopsis thalia
38	39	57.4	198	20	AAW18084	Histamine binding
39	39	57.4	264	21	AAW27731	Sequence homologou
40	39	57.4	285	19	AAW77301	Amino acid sequenc
41	39	57.4	379	21	AAW59390	Murine soluble int
42	39	57.4	392	21	AAW42988	Human ORFX ORF2752
43	39	57.4	432	17	AAW92813	Murine interleukin
44	39	57.4	432	22	AAW36653	Mouse IL-11 recept
45	39	57.4	441	17	AAW90991	Murine Etl-2 gene
46	39	57.4	454	21	AAW25433	Arabidopsis thalia
47	39	57.4	454	21	AAW34331	zea mays protein f
48	39	57.4	487	21	AAW17934	Arabidopsis thalia
49	39	57.4	495	21	AAW34330	zea mays protein f
50	39	57.4	500	21	AAW17933	Arabidopsis thalia
51	39	57.4	503	21	AAW25432	Arabidopsis thalia
52	39	57.4	521	20	AAW02367	Polypeptide ident
53	39	57.4	527	21	AAW34329	zea mays protein f
54	39	57.4	532	21	AAW17932	Arabidopsis thalia
55	39	57.4	535	21	AAW25431	Arabidopsis thalia
56	39	57.4	733	21	AAW52241	Arabidopsis thalia
57	39	57.4	779	21	AAW52240	Arabidopsis thalia
58	39	57.4	811	21	AAW52239	Arabidopsis thalia
59	39	57.4	1529	21	AAW96744	A. terreus ORF1 es
60	38	55.9	59	20	AAW41511	Fragment of human
61	38	55.9	77	21	AAW02476	Human secreted pro
62	38	55.9	190	17	AAW90768	FGF receptor ligan
63	38	55.9	331	21	AAW41847	Human ORFX ORF1611
64	38	55.9	336	21	AAW27990	Human secreted pro
65	38	55.9	336	22	AAW64536	Human secreted pro

■

CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergic rhinitis, sinusitis, bone and cartilage damage,
 CC nocturnal haemoglobinuria, anti-inflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 191 AA;
 SQ Query Match 66.2%; Score 45; DB 21; Length 191;
 Best Local Similarity 77.8%; Pred. No. 8; Mismatches 1; Indels 0; Gaps 0;
 Matches 7; Conservative

OY 2 GLLCPDPRP 10
 DB 9 gllcpdpap 17
 ||||| 1:1
 ||||| 1:1

RESULT 6
 AAY98874
 ID AAY98874 standard; Peptide: 15 AA.
 XX
 XX AAY98874;
 XX 07-AUG-2000 (first entry)
 XX HLA class II binding antigen epitope peptide #63.

XX Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;
 KW immune response; chronic viral disease; cancer; autoimmune disease;
 KW rheumatoid arthritis; allergic reaction; lyme disease; hepatitis; prostate cancer;
 KW allograft rejection; allergy; lyme disease; hepatitis; prostate cancer;
 KW glomerulonephritis; food hypersensitivity; malaria.

XX Unidentified.
 XX WO9961916-A1.
 XX 02-DEC-1999.
 XX 28-MAY-1999; 99WO-US12066.
 XX 29-MAY-1998; 98US-0087192.
 XX (EPIM-) EPIMUNE INC.
 XX Sette A, Southwood S, Sidney J;
 XX WPI; 2000-097143/08.

XX New compositions containing immunogenic peptide epitopes for various
 PT HLA class II DR molecules useful for inducing helper T cell response
 XX Claim 1; Page 41; 60pp; English.

XX The present invention relates to a new pharmaceutical composition
 CC comprising a unit dose form of a peptide or analogue of a peptide, an
 CC epitope derived from various antigens for various human leucocyte
 CC antigen class DR molecules, representative of the world wide population.
 CC The peptide/analogue binds to an HLA class II molecule at an IC-50 of
 CC less than or equal to 1,000 nM. The pharmaceutical can be used to induce
 CC a helper T cell response. The pharmaceutical focuses the immune response
 CC towards selected determinants and could therefore be used to treat
 CC chronic diseases and autoimmune diseases that can be
 CC treated using the peptide containing pharmaceutical include autoimmune
 CC diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia
 CC gravis), allograft rejection, allergies, lyme disease, hepatitis,
 CC post-streptococcal endocarditis or glomerulonephritis and food
 CC hypersensitivities. The peptide epitopes can be used to enhance immune
 CC responses against other immunogens administered mixtures include
 CC prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma,
 CC cervical carcinoma, lymphoma, and condyloma acuminatum. The peptides may

CC also be used to make monoclonal antibodies useful as potential diagnostic
 CC or therapeutic agents. The peptides may also be useful as diagnostic
 CC reagents, for example, to determine the susceptibility of an individual
 CC to a treatment regimen. Also, the peptides may be used to predict which
 CC individuals will be at substantial risk of developing chronic infection.
 CC The selection of appropriate T and B cell epitopes should allow the
 CC development of appropriate vaccines. The peptides may also be used to
 CC identify epitopes of antigens which are characterized by high sequence
 CC variability such as HIV, HCV and Malaria.

XX Sequence 15 AA;
 SQ Query Match 60.3%; Score 41; DB 21; Length 15;
 Best Local Similarity 66.7%; Pred. No. 4;
 Matches 6; Conservative

OY 2 GLLCPDPRP 10
 DB 2 gllcpdpap 10
 ||||| 1:1
 ||||| 1:1

RESULT 7
 AAY98970
 ID AAY98970 standard; Peptide: 15 AA.
 XX
 XX AAY98970;
 XX 07-AUG-2000 (first entry)
 XX HLA class II binding antigen epitope peptide #159.

XX Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;
 KW immune response; chronic viral disease; cancer; autoimmune disease;
 KW rheumatoid arthritis; allergic reaction; lyme disease; hepatitis; prostate cancer;
 KW allograft rejection; allergy; lyme disease; hepatitis; prostate cancer;
 KW glomerulonephritis; food hypersensitivity; malaria.

XX Unidentified.
 XX WO9961916-A1.
 XX 02-DEC-1999.
 XX 28-MAY-1999; 99WO-US12066.
 XX 29-MAY-1998; 98US-0087192.
 XX (EPIM-) EPIMUNE INC.
 XX Sette A, Southwood S, Sidney J;
 XX WPI; 2000-097143/08.

XX New compositions containing immunogenic peptide epitopes for various
 PT HLA class II DR molecules useful for inducing helper T cell response
 XX Claim 1; Page 42; 60pp; English.

XX The present invention relates to a new pharmaceutical composition
 CC comprising a unit dose form of a peptide or analogue of a peptide, an
 CC epitope derived from various antigens for various human leucocyte
 CC antigen class DR molecules, representative of the world wide population.
 CC The peptide/analogue binds to an HLA class II molecule at an IC-50 of
 CC less than or equal to 1,000 nM. The pharmaceutical can be used to induce
 CC a helper T cell response. The pharmaceutical focuses the immune response
 CC towards selected determinants and could therefore be used to treat
 CC chronic diseases and autoimmune diseases that can be
 CC treated using the peptide containing pharmaceutical include autoimmune
 CC diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia
 CC gravis), allograft rejection, allergies, lyme disease, hepatitis,
 CC post-streptococcal endocarditis or glomerulonephritis and food

CC hypersensitivities. The peptide epitopes can be used to enhance immune
 CC responses against other immunogens administered with the peptides.
 CC Diseases which can be treated using immunogenic mixtures include
 CC prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma,
 CC cervical carcinoma, lymphoma, and condyloma acuminatum. The peptides may
 CC also be used to make monoclonal antibodies useful as potential diagnostic
 CC or therapeutic agents. The peptides may also be useful as diagnostic
 CC reagents, for example, to determine the susceptibility of an individual
 CC to a treatment regimen. Also, the peptides may be used to predict which
 CC individuals will be at substantial risk of developing chronic infection.
 CC The selection of appropriate T and B cell epitopes should allow the
 CC development of epitope based vaccines particularly towards conserved
 CC epitopes of pathogens which are characterized by high sequence
 CC variability such as HIV, HCV and Malaria.
 XX
 XX Sequence 15 AA;

Query Match 60.3%; Score 41; DB 21; Length 15;
 Best Local Similarity 66.7%; Pred. No. 4;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLICPPRP 10
 | |||| |
 Db 2 gffcpdpap 10

RESULT 8
 AAB21202
 ID AAB21202 standard; protein; 59 AA.
 AC AAB21202;
 XX

DT 12-JAN-2001 (first entry)
 XX Human HER-2/neu protein phosphorylation domain partial sequence.

DE Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer.
 XX Homo sapiens.
 OS WO200044899-A1.
 PN 03-AUG-2000.
 XX

PF 28-JAN-2000; 2000WO-US02164.
 XX 29-JAN-1999; 99US-0117976.
 PR (CORI-) CORIXA CORP.
 PA (SMIK) SMITHKLINE BEECHAM.
 XX Cheever MA, Gheysen D;
 PI WPI; 2000-505976/45.
 DR

PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX Claim 27; Fig 11; 128pp; English.

CC The present sequence is a preferred portion of the phosphorylation domain
 CC of the HER-2/neu protein. HER-2/neu is a member of the tyrosine kinase
 CC family of receptor-like glycoproteins and shows homology to the epidermal
 CC growth factor receptor (EGFR). It probably plays a part in cell growth
 CC and/or differentiation. The HER-2/neu gene is an oncogene. An HER-2/neu
 CC fusion protein comprising a HER-2/neu extracellular domain fused to a
 CC HER-2/neu phosphorylation domain may be used to treat or prevent cancer
 CC by eliciting or enhancing an immune response to the HER-2/neu protein. It
 CC may be used to treat malignancies such as breast, ovarian, colon, lung

CC and prostate cancers, and may be used as an antigen to vaccinate against
 CC these neoplasias.
 CC
 CC Sequence 59 AA;
 Query Match 60.3%; Score 41; DB 21; Length 59;
 Best Local Similarity 66.7%; Pred. No. 14;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLICPPRP 10
 | |||| |
 Db 40 gffcpdpap 48

RESULT 9
 AAB21201
 ID AAB21201 standard; protein; 266 AA.
 AC AAB21201;
 XX

DT 12-JAN-2001 (first entry)
 XX Human HER-2/neu protein phosphorylation domain.

DE Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer.
 XX Homo sapiens.
 OS WO200044899-A1.
 PN 03-AUG-2000.
 XX

PF 28-JAN-2000; 2000WO-US02164.
 XX 29-JAN-1999; 99US-0117976.
 PR (CORI-) CORIXA CORP.
 PA (SMIK) SMITHKLINE BEECHAM.
 XX Cheever MA, Gheysen D;
 PI WPI; 2000-505976/45.
 DR

PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX Claim 2; Fig 10; 128pp; English.

CC The present sequence is the phosphorylation domain of the HER-2/neu
 CC protein. HER-2/neu is a member of the tyrosine kinase family of
 CC receptor-like glycoproteins and shows homology to the epidermal growth
 CC factor receptor (EGFR). It probably plays a part in cell growth and/or
 CC differentiation. The HER-2/neu gene is an oncogene. An HER-2/neu fusion
 CC protein comprising a HER-2/neu extracellular domain fused to a HER-2/neu
 CC phosphorylation domain may be used to treat or prevent cancer by
 CC eliciting or enhancing an immune response to the HER-2/neu protein. It
 CC may be used to treat malignancies such as breast, ovarian, colon, lung
 CC and prostate cancers, and may be used as an antigen to vaccinate against
 CC these neoplasias.
 XX Sequence 266 AA;

Query Match 60.3%; Score 41; DB 21; Length 266;
 Best Local Similarity 66.7%; Pred. No. 55;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLICPPRP 10
 | |||| |

Db 40 gffcpdpap 48

RESULT 10

AAW63682

ID AAW63682 standard; Protein: 291 AA.

XX

AC AAW63682;

AD

DT 24-SEP-1998 (first entry)

DE Human secreted protein 2.

XX

KW Secreted protein; human; cell proliferation; cytokine activity;

KW tissue growth; cellular differentiation; regeneration; activin;

KW tumour necrosis factor; tumour inhibition;

KW anti-inflammatory activity; biomarker.

XX

OS Homo sapiens.

PN WO9825959-A2.

PD 18-JUN-1998.

XX

PF 11-DEC-1997; 97WO-US22787.

PR 11-DEC-1996; 96US-0032757.

XX

XX (CHIR) CHIRON CORP.

XX

PI Escobedo J, Garcia P, Hu Q, Kothakota S, Williams LT;

DR WPI; 1998-348453/30.

DR N-PSDB; AAV43602.

XX

XX Secreted human polypeptides - having cytokine, cell proliferation or

XX differentiation activity or inhibin, tumour inhibition or

XX anti-inflammatory activities

PS Claim 1; Pages 49-50; 78pp; English.

XX

XX This represents a human secreted protein. The specification provides

XX a protein sequence, AAV43601, which is encoded by a cDNA

XX sequence showing a secreted polypeptide which is modified by rough

XX endoplasmic reticulum processing. The invention provides

XX a method of identifying a secreted polypeptide which is modified by rough

XX endoplasmic reticulum processing. The secreted proteins can be used in assays to determine

XX microsome. The secreted proteins can be used in assays to determine

XX differentiation activities, such as cytokine, cell proliferation, or cellular

XX inhibin activity, chemotactic or chemokinetic activity, haemostatic or

XX anti-inflammatory activity. The proteins can also be used as

XX biomarkers to identify tissues or cell types which express the proteins,

XX or a stage- or disease-specific alteration in protein expression. They

XX can be used in protein interaction assays, to identify ligands or binding

XX proteins. Compounds which affect the biological activities of the

XX secreted proteins or their ability to interact with specific ligands can

XX be used in assays to identify compounds which affect the biological

XX activities of the secreted proteins. The proteins can also be used to

XX antibodies that bind specifically to the protein can also be used to

XX design diagnostic tests and therapeutic compositions for diseases which

XX may be associated with altered expression of these proteins. Fusion

XX proteins comprising, e.g. signal sequences or transmembrane domains of

XX the proteins can be used to target other protein domains to cellular

XX membrane or they can be secreted extracellularly.

XX

SQ Sequence 291 AA;

Query Match 60.3%; Score 41; DB 19; Length 291;

Best Local Similarity 72.7%; Pred. No. 59;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GLICPDPRPLE 12

I I I I I I I I I I

Db 50 gqicwsprple 60

RESULT 11

AAW58875

ID AAW58875 standard; Protein: 293 AA.

XX

AC AAW58875;

AD

DT 27-MAR-2001 (first entry)

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 583.

XX

KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;

KW neoplastic; neuroprotective; antiviral; antitumor; anticonvulsant;

KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;

KW Addison's disease; allergy; autoimmune haemolytic anaemia;

KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;

KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;

KW cardiovascular disorder; wound healing; neurological disease.

XX

OS Homo sapiens.

PN WO200005173-A1.

PD 21-SEP-2000.

XX

PR 08-MAR-2000; 2000WO-US05881.

XX

XX 12-MAR-1999; 99US-0124270.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX Rosen CA, Ruben SM;

XX

XX WPI; 2000-611515/58.

XX

XX N-PSDB; AAF21778.

XX

XX New human breast and ovarian cancer associated gene sequences and the

XX polypeptides encoded by these genes, useful in the prevention,

XX treatment and diagnosis of cancer, immune disorders, cardiovascular

XX disorders and neurological diseases.

XX

XX Claim 11; Page 1019-1020; 1299pp; English.

XX

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human

XX proteins AAW58711 - AAW59128. The DNA and protein sequences are

XX associated with breast and ovarian cancer. The sequences included in the

XX invention are associated with breast and ovarian cancer. The sequences used in the

XX isolation and characterization of the DNA and protein sequences of the

XX invention. The breast and ovarian cancer associated DNA, protein, agonist

XX or antagonist sequences exhibit cytostatic; immunosuppressive;

XX neuroprotective; antiviral; antitumor; anticonvulsant;

XX antidiabetic; antiinflammatory; antitumor; antitumor; antitumor;

XX antibacterial; antifungal; antiparasitic; cardiant; immune disorder;

XX particularly breast and ovarian cancer. The nucleic acid sequences,

XX proteins, agonists and antagonists may also be used in the diagnosis,

XX prevention and treatment of immune disorders e.g. Addison's disease,

XX allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,

XX diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid

XX arthritis, ulcerative colitis, cardiovascular diseases such as

XX cerebral anoxia and epilepsy; and infectious diseases.

XX

SQ Sequence 293 AA;

Query Match 60.3%; Score 41; DB 21; Length 293;

Best Local Similarity 66.7%; Pred. No. 60;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GLLCPDPRP 10
 Db 113 gffcpdpap 121

RESULT 12

AA65496
 ID AAR65496 standard; Protein; 568 AA.
 XX
 AC AAR65496;
 XX
 DT 11-SEP-1995 (first entry)
 XX
 DE Marek's disease virus (MDV) protein.
 XX
 KW Marek's disease virus; MDV; antigenic proteins; vaccines;
 KW chickens; lymphocyte-growing diseases.
 XX
 OS Marek's disease virus.
 XX
 PN JP06292583-A.
 XX
 PD 21-OCT-1994.
 XX
 PF 16-MAR-1993; 93JP-0056066.
 XX
 PR 17-MAR-1992; 92US-0851971.
 PR 18-SEP-1992; 92US-0946231.
 XX
 PA (JAPG) NIPPON ZEON KK.
 PA (USGO) US GOVERNMENT.
 XX
 DR WPI: 1995-009074/02.
 DR N-PSDB; AAQ78645.
 XX

PT Antigenic protein originated from Marek's disease virus (MDV) -
 XX also recombinant MDV gene and a vaccine prepd. by using it
 XX
 PS Disclosure; Pages 8-17; 27pp; Japanese.
 XX

CC AAQ78645 encodes AAR65493, AAR65494, AAR65495; antigenic proteins for
 CC Marek's disease virus (MDV), and AAR65496. The antigenic proteins
 CC can be used in the development of an anti-MDV vaccine, for the
 CC treatment of infectious lymphocyte growing diseases in chickens,
 CC caused by the MDV virus.
 XX

SQ Sequence 568 AA;

Query Match 60.3%; Score 41; DB 16; Length 568;
 Best Local Similarity 54.5%; Pred. No. 1.1e+02;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 GLLCPDPRPLE 12
 Db 479 gvicpdpapad 489

RESULT 13

AAB21204
 ID AAB21204 standard; protein; 712 AA.
 XX
 AC AAB21204;
 XX
 DT 12-JAN-2001 (first entry)
 XX
 DE Human HER-2/neu fusion protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer; fusion protein.
 XX
 OS Homo sapiens.

OS Synthetic.
 XX WO200044899-A1.
 PN
 XX 03-AUG-2000.
 PD
 XX
 PF 28-JAN-2000; 2000WO-US02164.
 XX
 PR 29-JAN-1999; 99US-0117976.
 XX
 PA (CORI-) CORIXA CORP.
 PA (SMIK) SMITHLINE BEECHAM.
 XX
 PI Cheever MA, Gheysen D;
 XX
 DR WPI: 2000-505976/45.
 XX

PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX

PS Claim 27; Fig 13; 128pp; English.

XX The present sequence is a fusion protein comprising the extracellular
 CC domain and a preferred portion of the phosphorylation domain of the human
 CC HER-2/neu protein. HER-2/neu is a member of the tyrosine kinase family of
 CC receptor-like glycoproteins and shows homology to the epidermal growth
 CC factor receptor (EGFR). It probably plays a part in cell growth and/or
 CC differentiation. The HER-2/neu gene is an oncogene. HER-2/neu fusion
 CC proteins may be used to treat or prevent cancer by eliciting or enhancing
 CC an immune response to the HER-2/neu protein. They may be used to treat
 CC malignancies such as breast, ovarian, colon, lung and prostate cancers,
 CC and may be used as an antigen to vaccinate against these neoplasias.
 XX

SQ Sequence 712 AA;

Query Match 60.3%; Score 41; DB 21; Length 712;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GLLCPDPRP 10
 Db 693 gffcpdpap 701

RESULT 14

AAU00020
 ID AAU00020 standard; Protein; 729 AA.
 XX

AC AAU000020;

DT 09-MAY-2001 (first entry)

DE Human Plectin protein.

XX Human; Plectin; semaphorin domain; hyperplasia; neoplasia; cancer;
 KW neurodegenerative disease; autoimmune disease; lupus; multiple sclerosis;
 KW inflammatory bowel disease; diabetes type I; rheumatoid arthritis;
 KW immunogen; antibody.
 XX

OS Homo sapiens.

PN WO200114420-A2.

PD 01-MAR-2001.

PF 25-AUG-2000; 2000WO-US23365.

PR 25-AUG-1999; 99US-0150576.

PA (UYTO-) UNIV TORINO.

PA (REGC) UNIV CALIFORNIA.

PT		useful for vaccinating against breast, ovarian, colon, lung and prostate cancers -
XX		Claim 2; Fig 12; 128pp; English.
XX		The present sequence is a fusion protein comprising the extracellular domain and the phosphorylation domain of the human HER-2/neu receptor-like glycoprotein, and shows homology to the epidermal growth factor receptor (EGFR). It probably plays a part in cell growth and/or differentiation. The HER-2/neu gene is an oncogene. HER-2/neu fusion proteins may be used to treat or prevent cancer by eliciting or enhancing an immune response to the HER-2/neu protein. They may be used to treat malignancies such as breast, ovarian, colon, lung and prostate cancers, and may be used as an antigen to vaccinate against these neoplasias.
XX		Sequence 919 AA;
XX		
XX		Query Match 60.3%; Score 41; DB 21; Length 919;
XX		Best Local Similarity 66.7%; Pred. No. 1.7e+02;
XX		Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY	2	GILCCDDPPR 10
DB	693	gfccpdpap 701
RESULT	16	
AB21208		
AB21208		AASB21208 standard; Protein; 1200 AA.
AC	AAB21208;	
DT	12-JAN-2001	(first entry)
XX		Human HER-2/neu protein.
XX		Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine; breast cancer; prostate cancer; ovarian cancer; lung cancer; colon cancer.
OS	Homo sapiens.	
FN	WO200044899-A1.	
PA	03-AUG-2000.	
PD		
PF	28-JAN-2000; 2000WO-US02164.	
PR	29-JAN-1999; 99US-0117976.	
XX		(CORI-) CORIXA CORP.
PA	(SMIR) SMITHKLINE-BEECHAM.	
PI	Cheever MA, Gheysen D;	
XX		WPI; 2000-505976/45.
DR	N-PSDB; AAA89736.	
XX		HER-2/neu extracellular domain/phosphorylation domain fusion proteins useful for vaccinating against breast, ovarian, colon, lung and prostate cancers -
XX		Disclosure; Fig 15; 128pp; English.
XX		The present sequence is the human HER-2/neu protein. It is a member of the tyrosine kinase family of receptor-like glycoproteins and shows homology to the epidermal growth factor receptor (EGFR). It probably plays a part in cell growth and/or differentiation. The HER-2/neu gene is an oncogene. HER-2/neu fusion proteins comprising a HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation domain may be used to treat or prevent cancer by eliciting or

CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against
 CC these neoplasias.

SQ Sequence 1200 AA;

Query Match 60.3%; Score 41; DB 21; Length 1200;
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLICPDPRP 10
 | |||| |
 Db 1029 gffcpdpap 1037

RESULT 17

AAW01111
 ID AAW01111 standard; Protein; 1255 AA.

XX AC AAW01111;

XX DT 01-JAN-1997 (first entry)

XX DE HER-2/neu protein.

XX KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
 XX breast cancer; ovary cancer; colon cancer; lung cancer;
 XX prostate cancer; immunisation; tumour; vaccine; vector.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX FT Domain 676..1255

XX FT /label= Intracellular domain
 XX FT /note= "claimed domain, useful for immunisation"

XX PN W09630514-A1.

XX PD 03-OCT-1996.

XX PF 28-MAR-1996; 96WO-US01689.

XX PR 31-MAR-1995; 95US-0414417.

XX PA (UNIW) UNIV WASHINGTON.

XX PI Cheever MA, Disis ML;

XX DR WPI; 1996-455361/45.

XX DR N-PSDB; AAT40739.

XX PT DNA encoding HER-2/neu poly:peptide(s) - used for prevention or
 XX treatment of malignancies with which the HER-2/neu oncogene is
 XX associated

XX PS Claim 2; Page 56-61; 71pp; English.

XX CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
 CC the product of the HER-2/neu oncogene (see also AAT40739). The
 CC protein is over-expressed in various cancers, including breast,
 CC ovarian, colon, lung and prostate. The intracellular domain of the
 CC protein can be used to immunise an animal against a malignancy with
 CC which the oncogene is associated. The polypeptide can be produced
 CC in transformed host cells for use in immunisation. Alternatively,
 CC animal cells are transfected in vivo or ex vivo with a viral vector
 CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;

Query Match 60.3%; Score 41; DB 17; Length 1255;

Best Local Similarity 66.7%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLICPDPRP 10
 | |||| |
 Db 1029 gffcpdpap 1037

RESULT 18

AAW92406
 ID AAW92406 standard; Protein; 1255 AA.

XX AC AAW92406;

XX DT 21-APR-1999 (first entry)

XX DE Human HER-2/neu oncogene protein.

XX KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 XX malignancy; treatment; tumour.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX FT Region 676..1255

XX FT /note= "region which elicits immune response"

XX PN US5869445-A.

XX PD 09-FEB-1999.

XX PF 01-APR-1996; 96US-0625101.

XX PR 01-APR-1996; 96US-0625101.

XX PR 17-MAR-1993; 93US-0033644.

XX PR 12-AUG-1993; 93US-0106112.

XX PR 31-MAR-1995; 95US-0414417.

XX PA (UNIW) UNIV WASHINGTON.

XX PI Cheever MA, Disis ML;

XX DR WPI; 1999-152835/13.

XX DR N-PSDB; AAX01912.

XX PT Use of HER-2/neu polypeptides - for eliciting an immune response to
 XX an HER-2/neu associated malignancy, particularly for treating or
 XX preventing tumours

XX PS Claim 3; Column 31-38; 26pp; English.

XX CC This sequence represents the human HER-2/neu oncogene protein. A fragment
 CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.

XX SQ Sequence 1255 AA;

Query Match 60.3%; Score 41; DB 20; Length 1255;

Best Local Similarity 66.7%; Pred. No. 2.2e+02;

Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLICPDPRP 10
 | |||| |
 Db 1029 gffcpdpap 1037

RESULT 19

AAB21198

ID AAB211198 standard; protein; 1255 AA.
 AC AAB211198;
 XX
 DT 12-JAN-2001 (first entry)
 XX
 DE Human HER-2/neu protein.
 XX
 KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200044899-A1.
 XX
 PD 03-AUG-2000.
 XX
 PF 28-JAN-2000; 2000WO-US02164.
 XX
 PR 29-JAN-1999; 99US-0117976.
 XX
 RA (CORI-) CORIYA CORP.
 RA (SMK) SMITHKLINE BEECHAM.
 XX
 PI Cheever MA, Gheysen D;
 XX
 DR WPI; 2000-505976/45.
 DR N-PSDB; AAA89736.
 XX
 PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX
 PS Claim 52; Fig 7; 128pp; English.
 XX
 RA The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat cancer. The present protein may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against
 CC these neoplasias.
 CC
 XX
 SQ Sequence 1255 AA;
 Query Match 60.3%; Score 41; DB 21; Length 1255;
 Best Local Similarity 66.7%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 GLICPDPRP 10
 DB 1029 gffcdpdpap 1037
 RESULT 20
 AAY84780
 ID AAY84780 standard; Protein; 1255 AA.
 AC AAY84780;
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Amino acid sequence of the SPLICE erbB-2 receptor protein.
 XX
 KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
 KW tumor cell proliferation; tissue degeneration; arthropathy;
 KW bone resorption; inflammatory disease; degenerative disorder;

KW wound healing.
 XX
 OS Homo sapiens.
 XX
 PN WO200020579-A1.
 XX
 PD 13-APR-2000.
 XX
 XX 01-OCT-1999; 99WO-CA00912.
 XX
 PR 02-OCT-1998; 98US-0165192.
 XX
 RA (UYMC-) UNIV MCMASTER.
 XX
 XX Muller WJ, Siegel PM;
 XX WPI; 2000-303768/26.
 XX N-PSDB; AAA14812.
 XX
 PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
 PT erbB-2, inhibitors of the protein are useful for treatment of cancer -
 XX
 PS Claim 3; Fig 2; 60pp; English.
 XX
 RA The present sequence represents a SPLICE erbB-2 receptor protein. The
 CC protein has an in-frame deletion of 16 amino acids, 2 of which are
 CC conserved cysteine residues, compared to the unspliced protein. The
 CC erbB-2 polynucleotide is used to construct probes for detecting the
 CC protein. The probes may be used to detect SPLICE erbB-2 in a sample. Agents
 CC (e.g. antisense oligonucleotides) which inhibit the expression of
 CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
 CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
 CC for treating conditions of tissue degeneration, such as arthropathy, bone
 CC resorption, inflammatory disease, degenerative disorders of the
 CC central nervous system and wound healing.
 CC
 XX
 SQ Sequence 1255 AA;
 Query Match 60.3%; Score 41; DB 21; Length 1255;
 Best Local Similarity 66.7%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 GLICPDPRP 10
 DB 1029 gffcdpdpap 1037
 RESULT 21
 AAY92620
 ID AAY92620 standard; Protein; 1255 AA.
 AC AAY92620;
 XX
 XX 10-AUG-2000 (first entry)
 XX
 DE Human heregulin 2 (Her2).
 XX
 KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
 KW self-protein; cancer; breast cancer; prostate cancer;
 KW cell-associated peptide antigen; foreign epitope.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT Domain 1..173
 FT /label= "N-terminal
 FT /note= "mature polypeptide"
 FT Region 5055
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"

FT Region 59..73
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 103..117
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 149..163
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 174..323
 FT /label= Cysteine_rich_domain
 FT Region 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 324..483
 FT /label= Ligand_binding_domain
 FT Region 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 369..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 484..623
 FT /label= Cysteine_rich_domain
 FT Region 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 624..654
 FT /label= Transmembrane_domain
 FT Region 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 655..1010
 FT /label= Tyrosine_kinase_domain
 FT Region 661..675
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain 1011..1235
 FT /label= C-terminal_domain
 FT XX WO200020027-A2.
 XX 13-APR-2000.
 XX 05-OCT-1999; 99WO-DK00525.
 XX 05-OCT-1998; 98DK-0001261.
 XX 20-OCT-1998; 98US-0105011.
 XX (MEBI-) M & E BIOTECH AS.
 XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;
 PI Gautam A, Birk P, Karlsson G;
 XX WPI; 2000-349917/30.
 DR N-PSDB; AAA09455.
 XX Inducing immune responses to weakly immunogenic, tumor associated

PT peptide antigens for the treatment of breast and prostate cancer
 XX Claim 62; Page 193-198; 220pp; English.
 XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method
 CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animals immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 XX Sequence 1255 AA;
 SQ
 Query Match 60.3%; Score 41; DB 21; Length 1255;
 Best Local Similarity 66.7%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 2 GLLCPDPP 10
 Db 1029 gffcpdpap 1037
 RESULT 22
 AAB60167
 ID AAB60167 standard; Protein; 1255 AA.
 AC AAB60167;
 DT 03-APR-2001 (first entry)
 DX HER2 transgene plasmid construct encoded protein.
 DE Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
 KW antibody.
 KW Homo sapiens.
 OS Synthetic.
 OS OS
 PN WO200100244-A2.
 XX 04-JAN-2001.
 XX 23-JUN-2000; 2000WO-US17229.
 XX 25-JUN-1999; 99US-0141316.
 PR 16-MAR-2000; 2000US-0189844.
 XX (GETH) GENENTECH INC.
 PA Erickson S, Schwall R;
 PI WPI; 2001-061962/07.
 DR N-PSDB; AAF24297.
 XX Treating tumors, particularly breast cancers, which overexpress an ErbB
 PT receptor and does not respond to an anti-ErbB antibody, comprises
 PT conjugating the antibody to a maytansinoid -
 XX

PS Example 3; Fig 4; 92pp; English.

CC The present invention provides a method of treating cancer by
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
 CC particular, the antibody is directed against ErbB2 (also known as HER2
 CC and p185neu). The method is particularly useful in the treatment of
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
 XX Sequence 1255 AA;

Query Match 60.3%; Score 41; DB 22; Length 1255;
 Best Local Similarity 66.7%; Pred. No. 2.2e+02; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 3;
 QY 2 GLICPDPRP 10
 Db 1029 gffcpdpap 1037

RESULT 23
 AAR39568
 ID AAR39568 standard; Protein; 1433 AA.

XX AAR39568;
 XX 07-FEB-1994 (first entry)
 XX Sequence of c-erbB-2 tumour antigen.
 XX Tumour antigen; c-erbB-2; glycoprotein.

OS Homo sapiens.
 PA WO9316185-A.
 PI 19-AUG-1993.

XX 05-FEB-1993; 93WO-US01055.
 XX 06-FEB-1992; 92US-0831967.
 XX (CETU) CETUS ONCOLOGY CORP.
 XX (CREA.) CREATIVE BIOMOLECULES INC.

XX Houston LL, Huston JS, Oppermann H, Ring DB;
 XX WPI. 1993-273889/34.
 XX N-PSDB; AAQ46083.

XX New single chain Fv polypeptide binding to C-erbB-2 tumour
 XX antigen - for imaging or treating breast or ovarian cancer etc.
 XX Disclosure; pages 48-54; 87pp; English.

XX c-erbB-2 refers to a protein antigen expressed on the surface of
 XX tumour cells, such as breast and ovarian tumour cells, which is an
 XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
 XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents
 XX the location of a stop codon in AAQ46083.

XX Sequence 1433 AA;

Query Match 60.3%; Score 41; DB 14; Length 1433;
 Best Local Similarity 66.7%; Pred. No. 2.5e+02; Indels 0; Gaps 0;
 Matches 6; Conservative 0; Mismatches 3;

QY 2 GLICPDPRP 10
 Db 1029 gffcpdpap 1037

RESULT 24

AAU00019 standard; Protein; 2135 AA.

XX AAU00019;
 XX 09-MAY-2001 (first entry)
 XX Human plexin protein.

XX Human; Plexin; semaphorin domain; hyperplasia; neoplasia; cancer;
 XX retrodegeneration; diabetes mellitus; multiple sclerosis;
 XX inflammatory bowel disease; diabetes type I; rheumatoid arthritis;
 XX immunogen; antibody.

XX Homo sapiens.

XX Key Location/Qualifiers
 XX Key Location/Qualifiers
 XX Misc-difference /note= "Encoded by AAC"
 XX Misc-difference 1625
 XX /note= "Encoded by ACC"

XX WO200114420-A2.

XX 01-MAR-2001.

XX 25-AUG-2000; 2000WO-US23365.

XX 25-AUG-1999; 99US-0150576.

XX (UYTO-) UNIV TORINO.

XX (REGC) UNIV CALIFORNIA.

XX Artigiani S, Comoglio PM, Goodman CS, Tesier-Lavigne M;
 XX Tamagnone L;

XX WPI; 2001-226610/23.

XX N-PSDB; AAS00024.

XX New plexin polynucleotides and polypeptides, useful in diagnosis,
 XX therapy and in producing compounds for treating diseases involving
 XX aberrant cell growth (e.g. cancer) or immune regulation (e.g.
 XX autoimmune diseases)

XX Disclosure; Page 72-77; 79pp; English.

XX The sequence represents a Human Plexin. Plexins are large transmembrane
 XX proteins whose extracellular domain shares homology with Scatter factor
 XX receptors and contain an approximately 500 amino acid semaphorin domain.
 XX The plexin polynucleotides and polypeptides, and plexin-specific binding
 XX agents are useful in diagnosis, therapy and in the biopharmaceuticals
 XX industry for treating diseases involving aberrant cell growth (e.g. cancer)
 XX or in producing compounds (e.g. plexin-specific binding agents or
 XX antibodies) for treating or diagnosing a disease or disorder involving
 XX aberrant cell growth (e.g. hyperplasia, neoplasia, cancer or
 XX neurodegenerative disease), or diseases or disorders involving aberrant
 XX immune regulation (e.g. autoimmune diseases such as lupus, inflammatory
 XX bowel disease or diabetes Type I), or immunosuppressive diseases such as
 XX multiple sclerosis or rheumatoid arthritis.

XX Sequence 2135 AA;

Query Match 60.3%; Score 41; DB 22; Length 2135;
 Best Local Similarity 50.0%; Pred. No. 3.6e+02; Indels 0; Gaps 0;
 Matches 6; Conservative 3; Mismatches 3;

QY 1 AGLICPDPRPLE 12
 Db 584 sgvmcspdpse 595

RESULT 25
AA015105
ID AAR15105 standard; Protein; 122 AA.

XX AC AAR15105;
XX DT 11-FEB-1992 (first entry)
XX DE hCG/bLH chimera, D9.
XX KW Glycoprotein hormone; immuno-castration;
XX KW immuno-contraceptive; vaccine; human chorionic gonadotropin;
XX KW luteinising hormone; LH; CG; bovine.

XX OS Homo sapiens.
XX OS Bos taurus.

XX PN WO9116922-A.
XX PD 14-NOV-1991.

XX PF 07-MAY-1991; 91WO-US03162.

XX PR 08-MAY-1990; 90US-0520703.

XX PA (UYNE-) UNIV MED NEW JERSEY.

XX PI Campbell RK, Moyle WR;

XX DR WPI; 1991-353528/48.

XX PT New glyco-protein hormone analogues - for inducing fertility as
PT immuno-castration agents, for suppressing reproductive system
PT development and as immuno-contraceptive vaccines.

XX PS Table IV; Page 63; 94pp; English.

XX CC The sequence is an analogue of mature hCG beta subunit having
CC several residues replaced by the corresponding residues in the
CC bovine LH protein and residues 23-145 deleted. The chimeric hormone
CC may be useful for identifying residues which are important for
CC binding to the human receptor and may also have applications as an
CC immunogen, agonist and/or antagonist.
CC See AAR15043, AAR15061-R15125 and AAR15161-R15198.

XX SQ Sequence 122 AA;

Query Match 58.8%; Score 40; DB 12; Length 122;
Best Local Similarity 77.8%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPPDRPL 11
Db 108 itcdprpl 116

RESULT 26
AAW05504
ID AAW05504 standard; Protein; 328 AA.

XX AC AAW05504;

XX DT 15-JAN-1997 (first entry)

XX DE HCMV Toledo strain ULL135 protein (clone tol.03).

XX KW HCMV; vaccine; diagnosis; ULL135.

XX OS Human cytomegalovirus Toledo strain.

XX PN WO9630387-A1.

XX PD 03-OCT-1996.

XX PF 26-MAR-1996; 96WO-US04100.

XX PR 31-MAR-1995; 95US-0414926.

XX PA (AVIR-) AVIRON.

XX PI Cha T, Spaete R;

XX DR WPI; 1996-45265/45.

XX DR N-PSDB; AAT41418.

XX PT New isolated human cytomegalovirus nucleic acid - from Towne and
PT Toledo strains, used to develop prods. for the diagnosis, prevention
PT and treatment of human CMV infections

XX PS Claim 5; Page 69-70; 150pp; English.

XX CC Novel protein ULL135 (AAW05504) is the product of an open reading
CC frame found in a novel nucleic acid (AAT41418) isolated from the
CC Toledo strain of human cytomegalovirus (HCMV). ULL135 and other
CC novel (see also AAW05502-20) and known (see also AAW05500-01) proteins
CC of the Toledo strain, as well as new proteins (see also AAW05521-24)
CC from HCMV Towne, can be produced in transformed host cells and used
CC in the prodn. of subunit vaccines against HCMV. They may be
CC surface glycoproteins that are immunogenic or responsible for
CC tissue tropism, or may influence the immune response of an infected
CC individual.

XX SQ Sequence 328 AA;

Query Match 58.8%; Score 40; DB 17; Length 328;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10
Db 273 glscpcprp 281

RESULT 27

AAG31001
ID AAG31001 standard; Protein; 485 AA.

XX AC AAG31001;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 37160.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	19-APR-1999;	99US-0130057.	PR	19-JUL-1999;	99US-0144333.
PR	21-APR-1999;	99US-0130449.	PR	19-JUL-1999;	99US-0144335.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130831.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145086.
PR	04-MAY-1999;	99US-0132244.	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	99US-0132245.	PR	22-JUL-1999;	99US-0145087.
PR	05-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	11-MAY-1999;	99US-0132526.	PR	23-JUL-1999;	99US-0145145.
PR	12-MAY-1999;	99US-0132527.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	18-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	99US-0135233.	PR	28-JUL-1999;	99US-0145919.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146388.
PR	25-MAY-1999;	99US-0136031.	PR	02-AUG-1999;	99US-0146389.
PR	27-MAY-1999;	99US-0136392.	PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	99US-0137222.	PR	04-AUG-1999;	99US-0147302.
PR	02-JUN-1999;	99US-0137503.	PR	02-AUG-1999;	99US-0147162.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147303.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138847.	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0139119.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139452.	PR	11-AUG-1999;	99US-0148319.
PR	17-JUN-1999;	99US-0139493.	PR	12-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148564.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149729.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151062.
PR	22-JUN-1999;	99US-0139899.	PR	27-AUG-1999;	99US-0151062.
PR	23-JUN-1999;	99US-0140353.	PR	30-AUG-1999;	99US-0151303.
PR	24-JUN-1999;	99US-0140354.	PR	31-AUG-1999;	99US-0151438.
PR	24-JUN-1999;	99US-0140693.	PR	01-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	99US-0140823.	PR	07-SEP-1999;	99US-0152363.
PR	29-JUN-1999;	99US-0140929.	PR	10-SEP-1999;	99US-0153070.
PR	01-JUL-1999;	99US-0141842.	PR	11-SEP-1999;	99US-0153070.
PR	01-JUL-1999;	99US-0142154.	PR	15-SEP-1999;	99US-0154018.
PR	02-JUL-1999;	99US-0142055.	PR	16-SEP-1999;	99US-0154039.
PR	06-JUL-1999;	99US-0142390.	PR	20-SEP-1999;	99US-0154779.
PR	08-JUL-1999;	99US-0142803.	PR	22-SEP-1999;	99US-0155139.
PR	09-JUL-1999;	99US-0142920.	PR	23-SEP-1999;	99US-0155486.
PR	13-JUL-1999;	99US-0143542.	PR	24-SEP-1999;	99US-0155629.
PR	14-JUL-1999;	99US-0143624.	PR	29-SEP-1999;	99US-0156596.
PR	15-JUL-1999;	99US-0144005.	PR	04-OCT-1999;	99US-0157117.
PR	16-JUL-1999;	99US-0144085.	PR	05-OCT-1999;	99US-0157753.
PR	16-JUL-1999;	99US-0144086.	PR	06-OCT-1999;	99US-0157865.
PR	19-JUL-1999;	99US-0144325.	PR	07-OCT-1999;	99US-0158029.
PR	19-JUL-1999;	99US-0144325.	PR	08-OCT-1999;	99US-0158322.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159394.

```

PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0150329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      58.8%; Score 40; DB 21; Length 485;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LCPDPR 9
Db 375 lcpdpr 381

RESULT 28
AAG31000
ID AAG31000 standard; Protein; 486 AA.
XX
AC AAG31000;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37159.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.

```


CC their fragments, their nucleic acids and antibodies are used for
 CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
 CC infections, such as meningitis, septicaemia and gonorrhea. Both
 CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.

XX SQ Sequence 522 AA;

Query Match 58.8%; Score 40; DB 20; Length 522;
 Best Local Similarity 70.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLICPDRPL 11
 || |||||
 Db 48 glsvdpdrpv 57

RESULT 30
 AAY38568
 ID AAY38568 standard; Protein; 522 AA.

XX AC AAY38568;

DT 08-OCT-1999 (first entry)

DE *Neisseria* gonorrhoeae antigen encoded by ORF12.

XX *Neisseria* meningitidis; *Neisseria* gonorrhoeae; antigen; vaccine;
 KW treatment; *Neisseria* infection; meningitis; septicaemia; gonorrhea.
 XX *Neisseria* gonorrhoeae.

PN WO924578-A2.

XX 20-MAY-1999.

XX 09-OCT-1998; 98WO-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

XX (CHIR-) CHIRON SPA.

XX Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;

XX WPI; 1999-327407/27.

DR N-PSDB; AA212032.

XX Proteins from *Neisseria* meningitidis and *N. gonorrhoeae* useful for
 PT diagnosis, treatment and prevention of infection

XX Claim 4; Page 132; 524pp; English.

XX Amino acid sequences AAY38499-Y38944 represent *Neisseria* meningitidis
 CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open
 CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
 CC their fragments, their nucleic acids and antibodies are used for
 CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
 CC infections, such as meningitis, septicaemia and gonorrhea. Both
 CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.

XX SQ Sequence 522 AA;

Query Match 58.8%; Score 40; DB 20; Length 522;
 Best Local Similarity 70.0%; Pred. No. 1.4e+02;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 GLICPDRPL 11
 || |||||
 Db 48 glsvdpdrpv 57

RESULT 31

AAY38567

ID AAY38567 standard; Protein; 523 AA.

XX AC AAY38567;

DT 08-OCT-1999 (first entry)

DE *Neisseria* meningitidis strain A antigen encoded by ORF12.

XX *Neisseria* meningitidis; *Neisseria* gonorrhoeae; antigen; vaccine;
 KW treatment; *Neisseria* infection; meningitis; septicaemia; gonorrhea.
 XX *Neisseria* meningitidis.

PN WO924578-A2.

XX 20-MAY-1999.

XX 09-OCT-1998; 98WO-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

XX (CHIR-) CHIRON SPA.

XX Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;

XX WPI; 1999-327407/27.

DR N-PSDB; AA212031.

XX Proteins from *Neisseria* meningitidis and *N. gonorrhoeae* useful for
 PT diagnosis, treatment and prevention of infection

XX Claim 4; Page 130; 524pp; English.

XX Amino acid sequences AAY38499-Y38944 represent *Neisseria* meningitidis
 CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open
 CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
 CC their fragments, their nucleic acids and antibodies are used for
 CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
 CC infections, such as meningitis, septicaemia and gonorrhea. Both
 CC organisms are closely related. Fragments of the nucleic acids
 CC are useful as hybridisation probes and antisense reagents.

XX SQ Sequence 523 AA;

Query Match 58.8%; Score 40; DB 20; Length 523;
 Best Local Similarity 70.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLICPDRPL 11
 || |||||
 Db 49 glsvdpdrpv 58

RESULT 32

AAY74891

ID AAY74891 standard; Protein; 526 AA.

XX

AC AAY74891;
XX
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 302 protein sequence SEQ ID NO:1256.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHIR) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
XX N-PSDB; AA253653.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
XX Claim 2; Page 696; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisseria meningitidis and Neisseria gonorrhoeae, to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 526 AA;
XX
XX Query Match 58.8%; Score 40; DB 21; Length 526;
XX Best Local Similarity 70.0%; Pred. NO. 1.4e+02;
XX Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 2 GLICDPRPL 11
XX 11111111
XX Db 59 glsvdprrpv 68
XX
XX RESULT 33
XX AAY74890
XX AAAY74890 standard; Protein; 533 AA.
XX AC AAY74890;

XX
XX
XX 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 302 protein sequence SEQ ID NO:1254.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHIR) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
XX N-PSDB; AA253652.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics
XX
XX Claim 2; Page 695; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunogenic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisseria meningitidis and Neisseria gonorrhoeae, to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 533 AA;
XX
XX Query Match 58.8%; Score 40; DB 21; Length 533;
XX Best Local Similarity 70.0%; Pred. NO. 1.5e+02;
XX Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 2 GLICDPRPL 11
XX 11111111
XX Db 59 glsvdprrpv 68
XX
XX RESULT 34
XX AAY74892
XX AAAY74892 standard; Protein; 533 AA.
XX AC AAY74892;

DT 21-MAR-2000 (first entry)
XX Neisseria meningitidis ORF 302 protein sequence SEQ ID NO:1258.
DE
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzi M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
DR WPI; 2000-062150/05.
DR N-PSDB; AA253654.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 697-698; 1453pp; English.
XX
CC AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 533 AA;

Query Match 58.8%; Score 40; DB 21; Length 533;
Best Local Similarity 70.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRLL 11
|||
Db 59 gisvpdprpv 68

RESULT 35
AAG30999
ID AAG30999 standard; Protein; 538 AA.
XX
AC AAG30999;
XX
DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 37158.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.

OS	Zea mays subsp. may.	PR	30-JUN-1999;	99US-0141287.
PN		PR	01-JUL-1999;	99US-0141842.
XX	EP1033405-A2.	PR	01-JUL-1999;	99US-0142154.
XX		PR	02-JUL-1999;	99US-0142055.
PD	06-SEP-2000.	PR	06-JUL-1999;	99US-0142390.
XX		PR	08-JUL-1999;	99US-0142803.
XX		PR	09-JUL-1999;	99US-0142920.
PF	25-FEB-2000; 2000EP-0301439.	PR	12-JUL-1999;	99US-0142977.
PF		PR	13-JUL-1999;	99US-0143542.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	23-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145085.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145087.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145145.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	PR	03-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	17-JUN-1999;	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149902.
PR	21-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	21-JUN-1999;	PR	25-AUG-1999;	99US-0150566.
PR	23-JUN-1999;	PR	25-AUG-1999;	99US-0150884.
PR	23-JUN-1999;	PR	26-AUG-1999;	99US-0151065.
PR	24-JUN-1999;	PR	27-AUG-1999;	99US-0151066.
PR	24-JUN-1999;	PR	27-AUG-1999;	99US-0151080.
PR	24-JUN-1999;	PR	30-AUG-1999;	99US-0151303.
PR	24-JUN-1999;	PR	31-AUG-1999;	99US-0151438.
PR	28-JUN-1999;	PR	01-SEP-1999;	99US-0151930.
PR	28-JUN-1999;	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	PR	13-SEP-1999;	99US-0153758.


```
PR 12-JUL-1999; 99US-01429277.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 57.4%; Score 39; DB 21; Length 172;
Best Local Similarity 85.7%; Pred. No. 74;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9
Db 110 lvcppdr 116

RESULT 38
AAV18084
ID AAY18084 standard; Protein; 198 AA.
XX AC AAY18084;
XX DT 06-AUG-1999 (first entry)
XX DE Histamine binding protein AV-HBP.
XX Histamine binding protein; serotonin binding compound; inflammation;
KW gastric acid secretion; allergy; type I hypersensitivity reaction;
KW asthma; urticaria; allergic rhinitis; atopic dermatitis; food allergy;
KW drug allergy; abnormal blood pressure; psychological disorder; vaccine;
KW respiratory disease; coronary heart disease; cellular growth regulator;
KW tissue repair; blood-sucking ectoparasite; therapy.
XX Amblyomma variegatum.
XX OS
XX WO9927104-Al.
XX PN
XX PD 03-JUN-1999.
XX PF 26-NOV-1998; 98WO-GB03530.
XX PR 26-JUN-1998; 98GB-0013917.
XX PR 26-NOV-1997; 97GB-0025046.
```


PA (OXFO-) OXFORD VACS LTD.
 XX
 PI Nuttall PA, Paesen GC;
 XX
 DR WPI; 1999-357841/30.
 DR N-PSDB; AAX76969.
 XX
 XX Histamine and serotonin binding compounds useful for the treatment
 XX of allergies
 PT
 PS Claim 12; Fig 6; 84pp; English.
 XX
 CC This sequence is an example of a histamine or serotonin binding
 CC compound (A); of the invention. The compounds are useful for regulating
 CC acid secretion in the stomach, the detection, quantification and removal of gastric
 CC or serotonin (in animals, plants, cell cultures, food materials, or
 CC humans) and in the treatment of various diseases and allergies
 CC (e.g. type I hypersensitivity reactions, urticaria, asthma, allergic
 CC rhinitis (hay fever), atopic dermatitis, insect bites and food and drug
 CC allergies, abnormal blood pressure, migraine, psychomotor retardation,
 CC used to regulate cellular growth and tissue repair. The molecules may
 CC also be used as components of vaccines directed against blood-sucking
 CC ectoparasites.
 XX
 XX Sequence 198 AA;
 SQ
 Query Match 57.4%; Score 39; DB 20; Length 198;
 Best Local Similarity 85.7%; Pred. No. 84;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 4 ICDDPRP 10
 DB 177 ICDDPRP 183
 RESULT 39
 AAB27731
 ID AAB27731 standard; Protein: 264 AA.
 AC
 AC AAB27731;
 XX
 DT 30-JAN-2001 (first entry)
 XX
 DE Sequence homologous to protein fragment encoded by gene 2.
 XX
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 YN WO20005201-A1.
 PN
 XX 21-SEP-2000.
 PD
 XX 09-MAR-2000; 2000WO-US06059.
 PF
 XX 12-MAR-1999; 99US-0124096.
 PR 03-DEC-1999; 99US-0168622.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Rosen CA, Ruben SM, Komatsoulis G;
 PI
 XX WPI; 2000-628182/60.
 XX
 XX Novel human secreted proteins useful for diagnosis, prevention and
 PT treatment of disorders including neurological, cell proliferative,
 PT

PT cardiovascular, autoimmune/inflammatory disorders and microbial
 XX infections
 XX
 XX Disclosure: Page 9; 427pp; English.
 XX
 CC The invention relates to the isolation of genes AAC59157-C59205 encoding
 CC a human protein homologous to AAB27682-B27730. This sequence represents a
 CC peptide fragment homologous to the protein encoded by the gene given
 CC in the descriptor line. The sequence is a search result from a BLASTX
 CC homology search. The genes and proteins are useful for preventing,
 CC ameliorating or treating medical conditions, e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues and
 CC in the specification the number of genes is given. The genes are
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, d
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, disorders such
 CC as rheumatoid arthritis, Crohn's disease, multiple sclerosis, disorders such
 CC e.g. cerebral ischaemia; (d) wound healing; (e) neurological diseases
 CC as viral, bacterial, fungal and parasitic infections.
 XX
 XX Sequence 264 AA;
 SQ
 Query Match 57.4%; Score 39; DB 21; Length 264;
 Best Local Similarity 85.7%; Pred. No. 11e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 5 CPDPRPL 11
 DB 1 CPDPRPL 7
 RESULT 40
 AAW77301
 ID AAW77301 standard; Protein: 285 AA.
 XX
 XX AAW77301;
 XX
 DT 07-JAN-1999 (first entry)
 DE
 DE Amino acid sequence of murine HELA2.
 XX
 XX Serine protease; regulation; cell activity; viability; HELA2; ATC2;
 KW RCOM43; testis; fertility; suppressor; testicular germ cell cancer;
 KW seminoma; testis-specific expression; antitumor; sperm development;
 KW infertility; mouse.
 XX
 OS Mus sp.
 XX
 YN WO9836054-A1.
 XX
 PD 20-AUG-1998.
 XX
 XX 13-FEB-1998; 98WO-AU000085.
 PF
 XX 18-NOV-1997; 97AU-0000422.
 PR 13-FEB-1997; 97AU-0005101.
 XX
 XX (AMRA-) AMRAD OPERATIONS PTY LTD.
 PA
 XX Antalis TM, Hooper JD;
 PI
 XX WPI; 1998-480768/41.
 DR N-PSDB; AAV59132.
 XX
 XX New serine protease(s) and kinase involved in regulating cell
 PT activity and viability - particularly the testis-specific protease
 PT HELA2 used for modulation of fertility and as tumour suppressor
 XX

Example 14; Fig 18A; 167pp; English.

The present sequence represents the amino acid sequence of murine HELA2. Human HELA2 was isolated from HeLa cells. HELA2 has high homology to serine proteases. The protein is involved in or associated with regulation of cell activity and/or viability. Administration of recombinant HELA2 (also called testisin) is used to increase fertility. Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of testicular germ cell cancers (seminoma) and is also expressed in some non-testicular cancers (of colon, pancreas, prostate and ovary), so is a marker/potential therapeutic target for cancer. The promoter from the HELA2 gene is useful for testis-specific expression of other genes, e.g. for gene therapy or modulation of fertility. Drugs that block activity of HELA2 should have antitumour activity (other than in testis) while in testis recombinant HELA2 should stop growth of tumours and normalise sperm development (eliminating the need for orchidectomy). Identification of mutant forms of HELA2 can be used to diagnose infertility.

Sequence 285 AA;

Query Match 57.4%; Score 39; DB 19; Length 285;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLICPDPRPL 11.
| | | | | | |
Db 260 glirpdpvpl 269

RESULT 41

AAY59390
ID AAY59390 standard; Protein; 379 AA.

AC AAY59390;

DT 14-MAR-2000 (first entry)

DE Murine soluble interleukin-11 receptor.

Interleukin-11; IL-11; IL-11R; human; bone density disorder; gp130;
IL-11 tertiary complex; glycoprotein 130; postmenopausal bone loss;
bone resorption inhibitor; bone formation; therapy.

Mus sp.

WO9595608-A2.

25-NOV-1999.

PF 19-MAY-1999; 99WO-CA00516.

PR 19-MAY-1998; 98CA-2237915.

PA (HAMI-) HAMILTON CIVIC HOSPITAL RES DEV CORP.

Shaghnessy S, Austin RC;

WPI; 2000-062377/05.

N-PSDB; AA240400.

Inhibiting formation of a tertiary complex for the treatment of osteoporosis -

Example 3; Page 46-50; 61pp; English.

This sequence is the soluble mouse interleukin-11 receptor (IL-11R). The invention relates to a method of treating or alleviating the symptoms of a pathological condition in which bone density is decreased comprises inhibiting the formation of a tertiary complex of IL-11, IL-11 receptor and glycoprotein 130 (gp130) in a mammalian patient suffering from such a condition. The method is used to treat or alleviate the symptoms of a

pathological condition in which bone density is decreased, especially postmenopausal bone loss. The IL-11 binding peptide is useful in the purification of IL-11 or in depleting IL-11 from a solution. TRAP (tartrate-resistant acid phosphatase) and bone marrow formation assays can be used for the identification of IL-11 antagonists. The method not only inhibits bone resorption and hence bone loss, but also increases the process of bone formation to increase bone density.

Sequence 379 AA;

Query Match 57.4%; Score 39; DB 21; Length 379;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPL 12
| | | | | | |
Db 352 lqdprrpld 360

RESULT 42

AAB42988
ID AAB42988 standard; Protein; 392 AA.

AC AAB42988;

DT 08-FEB-2001 (first entry)

Human ORFX ORF2752 polypeptide sequence SEQ ID NO:5504.

Human; open reading frame; ORFX; detection; cytotostatic; hepatotropic; vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

Homo sapiens.

WO200058473-A2.

05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 03-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shimkets RA, Leach M;
WPI; 2000-602362/57.

N-PSDB; AAC77197.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -

Claim 11; Page 4686-4687; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX

XX	WPI: 1996-171612/17.
XX	N-PSDB; RAT17868.
XX	Nucleic acid encoding haemopoietin receptor containing conserved
PT	amino acid motif esp. IL-11 receptor alpha chain - used for
XX	developing IL-11 (ant)agonists
XX	Claim 6; Page 42-44; 87pp; English.
XX	The murine interleukin-11 (IL-11) receptor alpha chain Nrl
CC	(AF928413) was identified by expression of DNA (AAT17868). Isolated
CC	receptor mouse IL-11 and IL-11R1, gp130, and IL-11R2, and
CC	used to generate a high affinity IL-11 receptor. gp130 to generate a high
CC	affinity IL-11 receptor. IL-11 can be used to develop
CC	agonists or antagonists of therapeutic appln. or in the treatment
CC	or diagnosis of conditions involving a deficiency of IL-11,
CC	excess IL-11 or aberrant effects of normal endogenous IL-11
CC	levels.
XX	Sequence 432 AA:
SO	

Query Match	57.4%	Score 39;	DB 17;	Length 432;
Best Local Similarity	77.8%	Pred. No. 1.7e+02;		
Matches 7;	Conservative	1;	Mismatches 1;	Indels 0;
Gaps				
QY	4	LCPDPRPLE	12	
Db	352	lqpdprlpd	360	
RESULT 44				
AAB36653				
ID	AAB36653	standard;	432	AA.
XX	XX			
XX	XX	AAB36653;		
XX	XX			
DT	13-MAR-2001	(first entry)		
XX	XX			
DE	Mouse IL-11 receptor subunit alpha protein	SEQ ID NO:10.		
XX	XX			
XX	XX	DNAX cytokine receptor subunit; DCRS2; receptor protein;		
KW	KW	immunological disorder; diagnosis; detection; drug screening;		
XX	XX			
XX	XX			

mus sp.
XX
XX WO2000073451-A1.
XX
XX
XX
XX 07-DEC-2000.
XX
XX 30-MAY-2000; 2000WO-US14867.
XX
XX
XX 01-JUN-1999; 99US-0322913.
XX
XX (SCHE) SCHERING CORP.
XX
XX Dowling LM, Timans JC, Gorman DM, Kastelein RA, Bazan FJ;
XX WPI; 2001-061536/07.
XX
XX
XX Novel composition comprising DRAX cytokine receptor subunit polypeptide
XX useful for regulating immune system function and for treating
XX immunological disorders
XX
XX Disclosure: Page 13-15; 93pp: English.
XX
XX The present invention describes a composition (I) comprising a
XX recombinant DRAX cytokine receptor subunit-2 (DCRS2) polypeptide.
XX The DCRS2 polypeptide is useful for binding ligands and for preparing
XX compositions containing the same. The compositions are useful for cell
XX proliferation, for diagnostic and therapeutic applications, for cell
XX

CC detecting presence of their ligands and in drug screening assays. It
CC is also useful for treating conditions such as immunological disorders.
CC The present sequence represents a cytokine receptor subunit protein
CC which is given in an alignment of various cytokine receptor subunits in
CC the exemplification of the present invention.

XX
SQ Sequence 432 AA;

Query Match 57.4%; Score 39; DB 22; Length 432;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPLE 12
| | | | | | | |
Db 352 lqpdprpld 360

RESULT 45
AAR9091
ID AAR9091 standard; Protein; 441 AA.
XX AC AAR9091;
XX 09-OCT-1996 (first entry)
XX Murine Etl-2 gene product.
DE Interleukin-11 receptor; IL-11 receptor; antibody; antagonist; bone;
KW osteoporosis; Paget disease; myeloma; Etl-2.
XX OS Mus sp.
XX PN WO9619574-A1.
XX PD 27-JUN-1996.
XX PF 27-NOV-1995; 95WO-US15400.
XX PR 22-DEC-1994; 94US-0362304.
XX PA (GEMY) GENETICS INST INC.
XX PI Tobin JF;
XX WPI; 1996-309588/31.
XX DR N-PSDB; AAT32613.
XX PT New nucleic acid encoding human interleukin 11 receptor - and
XX related protein, antibodies, receptor antagonists, etc, useful for
XX treating and preventing loss of bone mass
PS Example 1; Page 37-40; 54pp; English.

CC The amino acid sequence of the murine Etl-2 gene product is given
CC in AAR9091. Probes based on the Etl-2 gene (AAT32613) were used
CC to screen a human cDNA library, yielding a cDNA clone (AAT33278) that
CC coded for human interleukin-11 receptor (AAR9090), a protein
CC involved in bone maturation and repair.

XX
SQ Sequence 441 AA;

Query Match 57.4%; Score 39; DB 17; Length 441;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPLE 12
| | | | | | | |
Db 361 lqpdprpld 369

Search completed: June 28, 2001, 11:54:14
Job time: 75 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:09 ; Search time 37.02 Seconds
(without alignments)
6.530 Million cell updates/sec

Title: US-09-439-313-558
Perfect score: 68
Sequence: 1 AGLLCPDRPLE 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Issued Patents_AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	63.2	362	2 US-09-055-097-7	Sequence 7, Appl
2	41	60.3	580	1 US-08-414-417B-69	Sequence 69, Appl
3	41	60.3	580	2 US-08-486-348A-69	Sequence 69, Appl
4	41	60.3	580	2 US-08-468-545B-69	Sequence 69, Appl
5	41	60.3	580	3 US-08-466-680B-69	Sequence 69, Appl
6	41	60.3	1255	1 US-08-467-083-68	Sequence 68, Appl
7	41	60.3	1255	1 US-08-414-417B-68	Sequence 68, Appl
8	41	60.3	1255	2 US-08-484-438-8	Sequence 8, Appl
9	41	60.3	1255	2 US-08-486-348A-68	Sequence 68, Appl
10	41	60.3	1255	2 US-08-625-101-2	Sequence 2, Appl
11	41	60.3	1255	2 US-08-468-545B-68	Sequence 68, Appl
12	41	60.3	1255	2 US-08-356-786-2	Sequence 2, Appl
13	41	60.3	1255	3 US-08-466-680B-68	Sequence 68, Appl
14	40	58.8	328	1 US-08-414-926A-9	Sequence 9, Appl
15	40	58.8	328	2 US-08-926-922-9	Sequence 9, Appl
16	40	58.8	328	3 US-09-253-682-9	Sequence 9, Appl
17	39.5	58.1	595	2 US-08-677-049-11	Sequence 11, Appl
18	38	55.9	190	1 US-08-441-629-4	Sequence 4, Appl
19	38	55.9	190	3 US-08-776-207-4	Sequence 4, Appl
20	38	55.9	190	5 PCT-US95-09172-4	Sequence 4, Appl
21	38	55.9	999	2 US-08-473-553A-5	Sequence 5, Appl
22	38	55.9	2595	4 US-09-036-987A-2	Sequence 2, Appl
23	37	54.4	279	1 US-08-300-903A-9	Sequence 9, Appl
24	37	54.4	617	4 US-09-188-930-303	Sequence 303, App
25	37	54.4	1290	1 US-08-470-350B-2	Sequence 2, Appl
26	37	54.4	3072	4 US-09-413-814-93	Sequence 93, Appl
27	37	54.4	3079	4 US-09-413-814-80	Sequence 80, Appl

28	36	52.9	96	1	US-08-414-926A-11	Sequence 11, Appl
29	36	52.9	96	2	US-08-926-922-11	Sequence 11, Appl
30	36	52.9	96	3	US-09-253-682-11	Sequence 11, Appl
31	36	52.9	233	1	US-08-221-750A-7	Sequence 7, Appl
32	36	52.9	1052	2	US-08-852-806-2	Sequence 2, Appl
33	36	52.9	1052	4	US-09-163-669-2	Sequence 2, Appl
34	35	51.5	288	3	US-08-934-494-4	Sequence 4, Appl
35	35	51.5	288	3	US-09-143-068-4	Sequence 4, Appl
36	35	51.5	732	1	US-08-317-522A-5	Sequence 5, Appl
37	35	51.5	778	1	US-08-439-818A-5	Sequence 5, Appl
38	35	51.5	778	2	US-08-751-965-5	Sequence 5, Appl
39	35	51.5	778	2	US-08-738-975-5	Sequence 5, Appl
40	35	51.5	778	2	US-08-728-626-5	Sequence 5, Appl
41	35	51.5	778	4	US-08-808-599A-5	Sequence 5, Appl
42	35	51.5	4472	2	US-08-804-227C-2	Sequence 2, Appl
43	34	50.0	60	2	US-08-465-794-14	Sequence 14, Appl
44	34	50.0	60	3	US-09-049-813-14	Sequence 14, Appl
45	34	50.0	80	4	US-08-663-191A-1	Sequence 1, Appl
46	34	50.0	80	4	US-09-051-624A-1	Sequence 1, Appl
47	34	50.0	95	1	US-07-891-942G-16	Sequence 16, Appl
48	34	50.0	157	2	US-08-599-602-2	Sequence 2, Appl
49	34	50.0	157	3	US-09-197-816-2	Sequence 2, Appl
50	34	50.0	178	2	US-08-465-794-18	Sequence 18, Appl
51	34	50.0	178	3	US-09-049-813-18	Sequence 18, Appl
52	34	50.0	178	4	US-08-663-191A-4	Sequence 4, Appl
53	34	50.0	189	4	US-09-369-494-15	Sequence 15, Appl
54	34	50.0	447	4	US-09-305-001-2	Sequence 2, Appl
55	34	50.0	521	2	US-08-504-048-9	Sequence 9, Appl
56	34	50.0	521	5	PCT-US93-08386-10	Sequence 10, Appl
57	34	50.0	548	3	US-08-656-034-10	Sequence 10, Appl
58	34	50.0	628	3	US-08-656-034-2	Sequence 2, Appl
59	34	50.0	1079	3	US-09-058-489-22	Sequence 22, Appl
60	34	50.0	1240	3	US-09-058-489-23	Sequence 23, Appl
61	34	50.0	1347	3	US-09-058-489-24	Sequence 24, Appl
62	33	48.5	51	4	US-08-905-223-384	Sequence 384, App
63	33	48.5	146	3	US-08-476-120-8	Sequence 8, Appl
64	33	48.5	225	1	US-08-300-903A-14	Sequence 14, Appl
65	33	48.5	234	1	US-08-300-903A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-055-097-7
; Sequence 7, Application US/09055097
; Patent No. 5955282
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: HUMAN OXIDIZED LDL RECEPTOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,097
; FILING DATE: Filed Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0490 US
; TELEPHONE: (650) 853-0555
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1465908
; US-09-055-097-7

Query Match          63.2%; Score 43; DB 2; Length 362;
Best Local Similarity 97.5%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPL 11
DB 92 LCPDPSPL 99

RESULT 2
US-08-414-417B-69
; Sequence 69, Application US/08414417B
; Patent No. 5801005
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Washington
; STATE: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; SOFTWARE: IBM PC compatible
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,417B
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C2
; TELEPHONE: (206) 682-6031
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-414-417B-69

Query Match          60.3%; Score 41; DB 1; Length 580;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10
DB 354 GFFCPDPAP 362

RESULT 4
US-08-486-348A-69
; Sequence 69, Application US/08486348A
; Patent No. 5876712
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; SOFTWARE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,348A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C6
; TELEPHONE: (206) 682-6031
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-486-348A-69

Query Match          60.3%; Score 41; DB 2; Length 580;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10
DB 354 GFFCPDPAP 362

RESULT 4
US-08-486-348B-69
; Sequence 69, Application US/08468545B
; Patent No. 5876712
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
```

STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/468.545B
APPLICATION NUMBER: US/08/468.545B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sharkey, Richard G.
REGISTRATION NUMBER: 32,629
REFERENCE/DOCKET NUMBER: 920010.448C5
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
US-08-468-545B-69

Query Match 60.3%; Score 41; DB 2; Length 580;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GLICPDPRP 10
| | | | |
Db 354 GFECDDPAP 362

RESULT 5
US-08-466-680B-69
Sequence 69, Application US/08466680B
Patent No. 6075122
GENERAL INFORMATION:
APPLICANT: Cheever, Martin A.
APPLICANT: Disis, Mary L.
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466.680B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sharkey, Richard G.
REGISTRATION NUMBER: 32,629
REFERENCE/DOCKET NUMBER: 920010.448C4
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
US-08-466-680B-69

Query Match 60.3%; Score 41; DB 3; Length 580;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GLICPDPRP 10
| | | | |
Db 354 GFECDDPAP 362

RESULT 6
US-08-467-083-68
Sequence 68, Application US/08467083
Patent No. 5726023
GENERAL INFORMATION:
APPLICANT: Cheever, Martin A.
APPLICANT: Disis, Mary L.
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
TITLE OF INVENTION: HER-2/NEU ONCOGENE IS ASSOCIATED
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,083
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/414,417
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Sharkey, Richard G.
REGISTRATION NUMBER: 32,629
REFERENCE/DOCKET NUMBER: 920010.448C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-467-083-68

Query Match 60.3%; Score 41; DB 1; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GLICPDPRP 10
| | | | |
Db 1029 GFECDDPAP 1037


```

RESULT 7
US-08-414-417B-68
; Sequence 68, Application US/08414417B
; Patent No. 5801005
; GENERAL INFORMATION:
; APPLICANT: Cheever, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414.417B
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Misstock, S. Leslie
; REGISTRATION NUMBER: 16,872
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 790-9090
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-414-417B-68

```

```

Query Match 60.3%; Score 41; DB 1; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 2 GLLCPDPRP 10
Db 1029 GFCCPDAP 1037

```

```

RESULT 8
US-08-484-438-8
; Sequence 8, Application US/08484438
; Patent No. 5811098 5780031
; GENERAL INFORMATION:
; APPLICANT: Plovman, Gregory D.
; APPLICANT: Shoyab, Mohammed
; APPLICANT: Siegal, Clay B.
; APPLICANT: Hellstr m, Ingegerd
; APPLICANT: Hellstr m, Karl E.
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: US
; ZIP: 10036-2711
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,438
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/323,442
; FILING DATE: 14-OCT-1994
; APPLICATION NUMBER: US 08/150,704
; FILING DATE: 10-NOV-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/981,165
; FILING DATE: 24-NOV-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misstock, S. Leslie
; REGISTRATION NUMBER: 16,872
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 790-9090
; LENGTH: 1255 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-484-438-8

```

```

Query Match 60.3%; Score 41; DB 2; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 2 GLLCPDPRP 10
Db 1029 GFCCPDAP 1037

```

```

RESULT 9
US-08-486-348A-68
; Sequence 68, Application US/08486348A
; Patent No. 5848638
; GENERAL INFORMATION:
; APPLICANT: Cheever, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,348A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-486-348A-68

Query Match 60.3%; Score 41; DB 2; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GLICPDPRP 10
| | | | |
Db 1029 GFFCPDPAP 1037

RESULT 10
US-08-625-101-2
; Sequence 2, Application US/08625101
; Patent No. 5869445
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE
; TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION
; TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu
; TITLE OF INVENTION: ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,101
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C7
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-625-101-2

Query Match 60.3%; Score 41; DB 2; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GLICPDPRP 10
| | | | |

Db 1029 GFFCPDPAP 1037

RESULT 11
US-08-468-545B-68
; Sequence 68, Application US/08468545B
; Patent No. 5876712
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,545B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-468-545B-68

Query Match 60.3%; Score 41; DB 2; Length 1255;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GLICPDPRP 10
| | | | |
Db 1029 GFFCPDPAP 1037

RESULT 12
US-08-356-786-2
; Sequence 2, Application US/08356786
; Patent No. 5877305
; GENERAL INFORMATION:
; APPLICANT: Huston, James S.
; APPLICANT: Oppermann, Hermann
; APPLICANT: Houston, L. L.
; APPLICANT: Ring, David B.
; TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer
; TITLE OF INVENTION: Marker
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault
; STREET: Exchange Place, 53 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356.786
FILING DATE: 07/831.967
CLASSIFICATION: 424
PRIORITY DATE: 07/831.967
APPLICATION NUMBER: 07/831.967
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-053
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-356-786-2

Query Match 60.3%; Score 41; DB 2; Length 1255;
Best Local Similarity 66.7%; Pred. No: 89;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLPCDPRP 10
DB 1029 GFPCDPRP 1037

RESULT 13
US-08-466-680B-68
; Sequence 68, Application US/08466680B
; Patent No. 6075122
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; ATTORNEY/AGENT INFORMATION:
; NAME: Cheever, Martin A.
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: tol.03
; FEATURE KEY: Protein
; LOCATION: 1-328
; OTHER INFORMATION: /label= UL135
US-08-414-926A-9
; Sequence 58.8%; Score 40; DB 1; Length 328;
; Best Local Similarity 77.8%; Pred. No: 34;
; Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLPCDPRP 10
DB 273 GLSCDPRP 281

RESULT 15
US-08-926-922-9
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/414.926A
FILING DATE: March 31, 1995
CLASSIFICATION: 435
PRIORITY DATE: March 31, 1995
APPLICATION NUMBER: US/08/414.926A
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: tol.03
FEATURE KEY: Protein
LOCATION: 1-328
OTHER INFORMATION: /label= UL135
US-08-414-926A-9
; Sequence 58.8%; Score 40; DB 1; Length 328;
; Best Local Similarity 77.8%; Pred. No: 34;
; Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLPCDPRP 10
DB 273 GLSCDPRP 281

RESULT 15
US-08-926-922-9
```

```
; Sequence 9, Application US/08926922
; Patent No. 5925751
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cseri Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,922
; FILING DATE: September 10, 1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.03
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..328
; OTHER INFORMATION: /label= UL135
;
US-08-926-922-9

Query Match 58.8%; Score 40; DB 2; Length 328;
Best Local Similarity 77.8%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPPDRP 10
DB 273 GLSCPCRP 281

RESULT 16
US-09-253-682-9
; Sequence 9, Application US/09253682
; Patent No. 6040170
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cseri Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/253,682
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/926,922
; FILING DATE: September 10, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 328 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.03
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..328
; OTHER INFORMATION: /label= UL135
;
US-09-253-682-9

Query Match 58.8%; Score 40; DB 3; Length 328;
Best Local Similarity 77.8%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPPDRP 10
DB 273 GLSCPCRP 281

RESULT 17
US-08-677-049-11
; Sequence 11, Application US/08677049
; Patent No. 5858707
; GENERAL INFORMATION:
; APPLICANT: Guimaraes, M. Jorge
; APPLICANT: Bazan, J. Fernando
; APPLICANT: McClanahan, Terrill K.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,049
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/000,788
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
```

REGISTRATION NUMBER: 34,090
REMARKS: PCT/US95/09172
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200
INFORMATION FOR SEQ ID NO: 11:
LENGTH: 595 amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 233..262
OTHER INFORMATION: /note= "Encompasses TM 4 of Figure 4"
FEATURE: Region
NAME/KEY: Region
LOCATION: 459..521
OTHER INFORMATION: /note= "Encompasses TM 9 and TM 10"
OTHER INFORMATION: of Figure 4"
US-08-677-049-11

Query Match 58.1%; Score 39.5; DB 2; Length 595;
Best Local Similarity 61.5%; Pred. No. 73;
Matches 8; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

QY 2 GLICPD--PRPL 11
DB 273 GMLCPSATAPRPL 285

RESULT 18
US-08-441-629-4
Sequence 4, Application US/08441629
Patent No. 5766923
GENERAL INFORMATION:
APPLICANT: Kirschner, Marc W.
APPLICANT: Kinoshita, No. 5766923iyuki
TITLE OF INVENTION: RECEPTOR-LIGAND ASSAY
FILE REFERENCE: HU95-01A2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: /08/441,629
FILING DATE: 15 MAY 1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER: US 08/279,217
FILING DATE: 22-JUL-1994
ATTORNEY/AGENT INFORMATION:
REGISTRATION NUMBER: 32,727
REFERENCE/DOCKET NUMBER: HU95-01A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 4:
LENGTH: 190 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-441-629-4
Query Match 55.9%; Score 38; DB 1; Length 190;
Best Local Similarity 72.7%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 3; Indels 0;
Gaps 0;
QY 2 GLICPDPRPLE 12
DB 28 GLNCNDPRLLE 38
RESULT 19
US-08-776-207-4
Sequence 4, Application US/08776207A
Patent No. 6080718
GENERAL INFORMATION:
APPLICANT: Kirschner, Marc W.
APPLICANT: Kinoshita, No. 6080718iyuki
TITLE OF INVENTION: RECEPTOR-LIGAND ASSAY
FILE REFERENCE: HU95-01A2
CURRENT APPLICATION NUMBER: US/08/776,207A
CURRENT FILING DATE: 1997-06-23
EARLIER APPLICATION NUMBER: PCT/US95/09172
EARLIER FILING DATE: 1995-07-19
EARLIER PUBLICATION NUMBER: 441,629
EARLIER FILING DATE: 1995-05-14
EARLIER APPLICATION NUMBER: 08/279,217
EARLIER FILING DATE: 1994-07-22
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 4
SEQ ID NO: 190
TYPE: PPT
ORGANISM: Xenopus laevis
US-08-776-207-4
Query Match 55.9%; Score 38; DB 3; Length 190;
Best Local Similarity 72.7%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 3; Indels 0;
Gaps 0;
QY 2 GLICPDPRPLE 12
DB 28 GLNCNDPRLLE 38

RESULT 20
PCT-US95-09172-4
Sequence 4, Application PC/TUS9509172
GENERAL INFORMATION:
APPLICANT: Kirschner, Marc W.
APPLICANT: Kinoshita, No. 5766923iyuki
TITLE OF INVENTION: RECEPTOR-LIGAND ASSAY
FILE REFERENCE: HU95-01A2
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: /08/441,629
FILING DATE: 15 MAY 1995
APPLICATION NUMBER: PCT/US95/09172
FILING DATE:

```
;
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/279,217
; FILING DATE: 22-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/441,629
; FILING DATE: 15-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: H095-01A PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-09172-4

Query Match 55.9%; Score 38; DB 5; Length 190;
Best Local Similarity 72.7%; Pred. No. 41;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GLLCPDPRPLE 12
   || | ||| ||
Db 28 GLNCNDPRLE 38

RESULT 21
US-08-473-553A-5
; Sequence 5, Application US/08473553A
; Patent No. 5859338
; GENERAL INFORMATION:
; APPLICANT: Meyerowitz, Elliot M.
; APPLICANT: Clark, Steven E.
; APPLICANT: Williams, Robert W.
; TITLE OF INVENTION: Plant Clavatal Nucleic Acids,
; TITLE OF INVENTION: Transformed Plants, and Proteins
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,553A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-60886/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 999 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown

;
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-473-553A-5

Query Match 55.9%; Score 38; DB 2; Length 999;
Best Local Similarity 70.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GLLCPDPRPL 11
   |||| | ||
Db 943 GLLCTSPPL 952

RESULT 22
US-09-036-987A-2
; Sequence 2, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-036-987A-2

Query Match 55.9%; Score 38; DB 4; Length 2595;
Best Local Similarity 72.7%; Pred. No. 5.4e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AGLLCPDPRPL 11
   |||| ||||
Db 2390 AGLLAARPL 2400

RESULT 23
US-08-300-903A-9
```



```
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 93
; LENGTH: 3072
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-93

Query Match          54.4%; Score 37; DB 4; Length 3072;
Best Local Similarity 66.7%; Pred. No. 9.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLICPDR 9
   ||: || ||
Db 44 AGVSCSPR 52

RESULT 27
US-09-413-814-80
; Sequence 80, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloecker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 3079
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-80

Query Match          54.4%; Score 37; DB 4; Length 3079;
Best Local Similarity 66.7%; Pred. No. 9.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLICPDR 9
   ||: || ||
Db 47 AGVSCSPR 55

RESULT 28
```

```
US-08-414-926A-11
; Sequence 11, Application US/08414926A
; Patent No. 5721354
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-2155
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414,926A
; FILING DATE: March 31, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-494-7622
; TELEFAX: 415-857-0663
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: tol.05
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..96
; OTHER INFORMATION: /label= ULL37
US-08-414-926A-11

Query Match          52.9%; Score 36; DB 1; Length 96;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLCPPDRP 10
   :||| ||
Db 26 VLCPLRP 33

RESULT 29
US-08-926-922-11
; Sequence 11, Application US/08926922
; Patent No. 5925751
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Cha, Tai-An
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Luann Cseri Attorney at Law
; STREET: 750 Arimo Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```



```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/926,922
; FILING DATE: September 10, 1997
; CLASSIFICATION: 536
; ATORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810
; INFORMATION:
; SEQUENCE CHARACTERISTICS: 11:
; LENGTH: 96 amino acids
; TYPE: amino acid
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: col.05
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..96
; OTHER INFORMATION: /label= UL137
;
; US-08-926-922-11
;
; Query Match 52.9%; Score 36; DB 2; Length 96;
; Best Local Similarity 75.0%; Pred. No. 43;
; Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
;
; QY 3 LLCPPDRP 10
; DB 26 VLCFDLRP 33
;
; RESULT 30
; Sequence 11, Application US/09253682
; Patent No. 43,917,000
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; SEELEE, Luann, Cseri Attorney at Law
; STREET: 750 Arlino Avenue
; CITY: Oakland
; STATE: CA
; COUNTRY: USA
; ZIP: 94610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/253,682
; FILING DATE:
; CLASSIFICATION:
; INFORMATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/926,922
; FILING DATE: September 10, 1997
; ATORNEY/AGENT INFORMATION:
; NAME: Cseri, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR 11A
; TELEPHONE: 510-834-1448
; TELEFAX: 510-839-7810

```

```

; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: col.05
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..96
; OTHER INFORMATION: /label= UL137
;
; US-08-253-682-11
;
; Query Match 52.9%; Score 36; DB 3; Length 96;
; Best Local Similarity 75.0%; Pred. No. 43;
; Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
;
; QY 3 LLCPPDRP 10
; DB 26 VLCFDLRP 33
;
; RESULT 31
; Sequence 7, Application US/08221750A
; Patent No. 5,507,772
; GENERAL INFORMATION:
; APPLICANT: Baker, Steven M.
; APPLICANT: Deich, Robert A.
; TITLE OF INVENTION: Genes for the Export of Pertussis
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 01774
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,750A
; FILING DATE: 31-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/031,619
; FILING DATE: 15-MAR-1993
; ATORNEY/AGENT INFORMATION:
; NAME: Hamilton, Brook, Smith & Reynolds, P.C.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: ACC93-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-9540
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-221-750A-7

```

```

; Query Match 52.9%; Score 36; DB 1; Length 233;
; Best Local Similarity 100.0%; Pred. No. 1e+02;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 6 PDRPL 11
111111
Db 2 PDRPL 7

RESULT 32

US-08-852-806-2
; Sequence 2, Application US/08852806
; Patent No. 5874245
; GENERAL INFORMATION:
; APPLICANT: Shoji Fukusumi
; APPLICANT: Shuji Hinuma
; TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,806
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017,915
; FILING DATE: 16 MAY 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: TAK50002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1052 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
US-08-852-806-2

Query Match 52.9%; Score 36; DB 2; Length 1052;
Best Local Similarity 66.7%; Pred. NO. 4.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLICPPRP 10
111111
Db 325 GLILPPRP 333

RESULT 33

US-09-163-669-2
; Sequence 2, Application US/09163669
; Patent No. 6111076
; GENERAL INFORMATION:
; APPLICANT: FUKUSUMI, SHOJI
; APPLICANT: HINUMA, SHUJI
; APPLICANT: FUJII, RIO
; TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED

; TITLE OF INVENTION: RECEPTOR (HIBCD07)
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/163,669
; FILING DATE: 30-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/852,806
; FILING DATE: 07-MAY-1997
; APPLICATION NUMBER: 60/017,915
; FILING DATE: 16-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: TAK-50002-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1052 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-163-669-2

Query Match 52.9%; Score 36; DB 4; Length 1052;
Best Local Similarity 66.7%; Pred. No. 4.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLICPPRP 10
111111
Db 325 GLILPPRP 333

RESULT 34

US-08-934-494-4
; Sequence 4, Application US/08934494
; Patent No. 6030831
; GENERAL INFORMATION:
; APPLICANT: Gurney, Austin
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,494
; FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1078
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQUENCES: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-934-494-4

Query Match 51.5%; Score 35; DB 3; Length 288;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CPDPRPLE 12
DB 29 CPGPRELE 36

RESULT 35
US-09-143-068-4
Sequence 4, 601873
GENERAL INFORMATION:
APPLICANT: Fong, Sherman
APPLICANT: Ferrara, Napoleone
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Guiney, Austin L.
APPLICANT: Hirsch, Michael
APPLICANT: Williams, Mickey
TITLE OF INVENTION: Tie Ligand Homologues
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
OPERATING SYSTEM: PC DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
FILING DATE: 28-Aug-1998
CLASSIFICATION INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1078P1
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQUENCES: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-09-143-068-4

Query Match 51.5%; Score 35; DB 3; Length 288;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 CPDPRPLE 12
DB 29 CPGPRELE 36

RESULT 36
US-08-317-522A-5
Sequence 5, Application US/08317522A
Patent No. 5599918
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
Proteins
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
CURRENT APPLICATION DATA:
FILING DATE: 04-OCT-1994
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REFERENCE/DOCKET NUMBER: P-LA 9991
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 72 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-522A-5

Query Match 51.5%; Score 35; DB 1; Length 732;
Best Local Similarity 70.0%; Pred. No. 4.5e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12
DB 581 CPGAEPREPLE 590

RESULT 37
US-08-439-818A-5
Sequence 5, Application US/08439818A
Patent No. 5654145
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
Proteins
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,818A
; FILING DATE: 12-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/317,522
; FILING DATE: 04-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 1563
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 778 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-439-818A-5

Query Match          51.5%; Score 35; DB 1; Length 778;
Best Local Similarity 70.0%; Pred. No. 4.8e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12
   || :|||||
Db 581 CPPAEPRPLE 590

RESULT 38
US-08-751-965-5
; Sequence 5, Application US/08751965
; Patent No. 5858360
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,965
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 12-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 778 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-751-965-5

Query Match          51.5%; Score 35; DB 2; Length 778;
Best Local Similarity 70.0%; Pred. No. 4.8e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12
   || :|||||
Db 581 CPPAEPRPLE 590

RESULT 39
US-08-738-975-5
; Sequence 5, Application US/08738975
; Patent No. 5880267
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,975
; FILING DATE: herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 05-Dec-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 778 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-738-975-5

Query Match          51.5%; Score 35; DB 2; Length 778;
Best Local Similarity 70.0%; Pred. No. 4.8e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12
   || :|||||
Db 581 CPPAEPRPLE 590

RESULT 40
US-08-728-626-5
; Sequence 5, Application US/08728626
; Patent No. 5910451
```

```

; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Proteins
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,626
; FILING DATE: 04-OCT-1994
; PRIORITY INFORMATION:
; PRIOR APPLICATION NUMBER: US 08/439,818
; FILING DATE: 12-MAY-1995
; APPLICATION DATA:
; APPLICANT: Fukuda, Michiko N.
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 1563
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 778 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-728-626-5

Query Match 51.5%; Score 35; DB 2; Length 778;
Best Local Similarity 70.0%; Pred. No. 4.8e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 1;

QY 5 CP--DPRPLE 12
Db 581 CPPAEPRPLE 590

RESULT 41
US-08-808-599A-5
; Sequence 5, Application US/08808599A
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin, Trophinin-Assisting
; TITLE OF INVENTION: Proteins and Methods to Inhibit Implantation
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/808,599A
; FILING DATE: 28-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/317,522
; FILING DATE: 04-OCT-1994
; PRIOR APPLICATION DATA: US 08/439,818
; FILING DATE: 12-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2256
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-8949
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 778 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; US-08-808-599A-5

Query Match 51.5%; Score 35; DB 4; Length 778;
Best Local Similarity 70.0%; Pred. No. 4.8e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

QY 5 CP--DPRPLE 12
Db 581 CPPAEPRPLE 590

RESULT 42
US-08-804-227C-2
; Sequence 2, Application US/08804227C
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuntz, Stuart A.
; APPLICANT: Kuntz, Stuart A., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: 15114 CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 41
; PRIOR APPLICATION INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4472 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

```

US-08-804-227C-2

Query Match 51.5%; Score 35; DB 2; Length 4472;
Best Local Similarity 66.7%; Pred. No. 2.7e+03;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 GLICPDPRP 10
| | | | |
Db 3638 GTLVDPDP 3646

RESULT 43

US-08-465-794-14
; Sequence 14, Application US/08465794
; Patent No. 5886141
; GENERAL INFORMATION:
; APPLICANT: FOLKMAN, MOSES J.
; APPLICANT: SHING, YUEN
; APPLICANT: IGARASHI, KOICHI
; TITLE OF INVENTION: SMOOTH MUSCLE MITOGEN AND ISOLATED DNA
; TITLE OF INVENTION: CODING THEREFORE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,794
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,126
; FILING DATE: 21-JAN-1993
; APPLICATION NUMBER: US 07/994,776
; FILING DATE: 22-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,597
; FILING DATE: 23-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,792
; FILING DATE: 23-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/833,552
; FILING DATE: 10-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/832,939
; FILING DATE: 10-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/766,354
; FILING DATE: 26-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/604,778
; FILING DATE: 26-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 40435-CIP-8
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 60 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-465-794-14

Query Match 50.0%; Score 34; DB 2; Length 60;
Best Local Similarity 85.7%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GLICPD 8
| | | | |
Db 13 GLICGDP 19

RESULT 44

US-09-049-813-14
; Sequence 14, Application US/09049813
; Patent No. 6013762
; GENERAL INFORMATION:
; APPLICANT: FOLKMAN, MOSES J.
; APPLICANT: SHING, YUEN
; APPLICANT: IGARASHI, KOICHI
; TITLE OF INVENTION: SMOOTH MUSCLE MITOGEN AND ISOLATED DNA
; TITLE OF INVENTION: CODING THEREFORE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/049,813
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/465,794
; FILING DATE:
; APPLICATION NUMBER: US 08/007,126
; FILING DATE: 21-JAN-1993
; APPLICATION NUMBER: US 07/994,776
; FILING DATE: 22-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,597
; FILING DATE: 23-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,792
; FILING DATE: 23-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/833,552
; FILING DATE: 10-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/832,939
; FILING DATE: 10-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/766,354
; FILING DATE: 26-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/604,778
; FILING DATE: 26-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNICK, DAVID S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 40435-CIP-8
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-3440
TELETYPE: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-049-813-14

Query Match 50.0%; Score 34; DB 3; Length 60;
Best Local Similarity 85.7%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GLICDPP 8
DB 13 GLICGDP 19

RESULT 45
US-08-663-191A-1
US-08-663-191A-1 Application US/08663191A
Patent No. 6183971
GENERAL INFORMATION:
APPLICANT: Reiko SASADA, et al.
TITLE OF INVENTION: ANTIBODY, HYBRIDOMA AND USE THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDING ADDRESS: 2033 K Street, N.W., Suite 800
ADDRESS: 2033 K Street, N.W., Suite 800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER: IBM compatible
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/663,191A
FILING DATE: 11-JUN-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Lee Cheng
REGISTRATION NUMBER: 40,949
PRACTICE ADDRESS: 36-0256/LC(WMC)/927
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 80 amino acid residues
STRANDNESS: N/A
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-08-663-191A-1

Query Match 50.0%; Score 34; DB 4; Length 80;
Best Local Similarity 85.7%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GLICDPP 8
DB 12 GLICGDP 18

Search completed: June 28, 2001, 11:56:11
Job time: 192 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:57 ; Search time 42.81 seconds
(without alignments)
21.352 Million cell updates/sec

Title: US-09-439-313-558
Perfect score: 68
Sequence: 1 AGLCPDPRPLE 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : PIR_68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	67.6	567	2 F64058	nitrate/nitrite se
2	44	64.7	397	2 H70753	hypothetical prote
3	44	64.7	357	2 T34362	hypothetical prote
4	41	60.3	362	1 C70646	probable pflA prot
5	41	60.3	397	2 B70815	probable transamin
6	41	60.3	430	2 B70813	hypothetical prote
7	41	60.3	568	2 JQ2206	UL46h protein - Ma
8	41	60.3	607	2 S01939	hypothetical prote
9	41	60.3	1255	1 A24571	protein-tyrosine k
10	40	58.8	140	2 G83562	hypothetical prote
11	40	58.8	210	2 G84499	hypothetical prote
12	40	58.8	261	2 D70871	hypothetical prote
13	40	58.8	276	2 G47301	VirB8 homolog - Bo
14	40	58.8	398	2 D85942	probable dehydrata
15	40	58.8	398	2 G65070	hypothetical prote
16	40	58.8	522	2 C81826	hypothetical integ
17	40	58.8	528	2 T52101	probable nuclear t
18	40	58.8	538	2 F86225	hypothetical prote
19	40	58.8	633	2 E75277	ABC transporter, A
20	39.5	58.1	399	2 D86177	protein Flp19.3 f
21	39.5	58.1	595	2 A48878	uric acid/xanthine
22	39	57.4	146	2 T53030	submaxillary prote
23	39	57.4	432	2 I48343	interleukin-11 rec
24	39	57.4	527	2 T04329	importin alpha - t
25	39	57.4	532	2 T52102	probable nuclear t
26	39	57.4	532	2 T52268	importin alpha [va
27	39	57.4	535	2 T52098	probable nuclear t
28	39	57.4	1207	1 BGHU	epidermal growth f
29	38.5	56.6	529	2 T35966	probable secreted

30	38	55.9	338	2	G72241	electron transfer
31	38	55.9	350	2	G75571	MutY/nudix family
32	38	55.9	389	2	S74343	probable aspartate
33	38	55.9	475	2	T33943	hypothetical prote
34	38	55.9	483	2	S74637	lysine decarboxyla
35	38	55.9	578	2	E70645	hypothetical prote
36	38	55.9	999	1	S27756	receptor-like prot
37	38	55.9	1264	2	S64146	probable membrane
38	38	55.9	1321	2	JE0352	mucin MUC5B, trach
39	38	55.9	1890	2	T35694	ATP dependent DNA
40	38	55.9	2139	2	T18296	myosin heavy chain
41	37	54.4	121	2	B83432	conserved hypotet
42	37	54.4	246	2	S63639	hypothetical prote
43	37	54.4	269	2	H85408	GATA transcription
44	37	54.4	269	2	T05288	GATA-binding trans
45	37	54.4	415	2	S28088	gene B protein - y
46	37	54.4	518	2	S61920	B-alpha pheromone-
47	37	54.4	629	2	T23538	hypothetical prote
48	37	54.4	649	2	T04606	protein kinase hom
49	37	54.4	653	2	G70683	probable lepA - My
50	37	54.4	659	2	S30893	nrtC protein - Syn
51	37	54.4	779	2	T31732	hypothetical prote
52	37	54.4	935	2	E96806	hypothetical prote
53	37	54.4	1290	2	A57190	ebnerin precursor
54	36	52.9	116	2	T34800	hypothetical prote
55	36	52.9	141	2	S67149	hypothetical prote
56	36	52.9	174	2	B82747	hypothetical prote
57	36	52.9	180	2	T12493	hypothetical prote
58	36	52.9	229	2	C71493	probable uracil DN
59	36	52.9	249	2	S76255	hypothetical prote
60	36	52.9	251	1	G71366	probable uridylyate
61	36	52.9	267	2	T36675	probable integral
62	36	52.9	291	1	JN0064	insulin-like growt
63	36	52.9	310	2	JN0490	28K lipase precurs
64	36	52.9	332	2	H83924	tryptophanyl-trna
65	36	52.9	352	2	I40177	maleylacetate redu

ALIGNMENTS

RESULT 1
F64058
nitrate/nitrite sensor protein (EC 2.7.3.-) [similarity] - Haemophilus influenzae (st
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 21-Jul-2000
C:Accession: F64058
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: F64058
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-567 <TIGR>
A:Cross-references: GB:U32713; GB:L42023; NID:g1573231; PIDN:AAC21933.1; PID:g1573233
C:Genetics:
A:Start codon: GTG
C:Superfamily: nitrate/nitrite sensor protein narX
C:Keywords: autophosphorylation; phosphohistidine; phosphoprotein; phosphotransferase
F:373/Active site: His (phosphohistidine intermediate) #status predicted

Query Match 67.6%; Score 46; DB 2; Length 567;
Best Local Similarity 66.7%; Pred. No. 6.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLCPDPRPLE 12

DB 326 AGLCPDPRIMQ 337

C/Accession: B70813
R/Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
Jr.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
K.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 53-54 (1998)
Nucleic Acids Res. 26, 1981-1988
A/Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A/Reference number: A70500; MUID: 98295987
A/Accession: B70813

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-430 <COL>
 C:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAA17648.1; PID:el25398
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: Rv0842

Query Match 60.3%; Score 41; DB 2; Length 430;
 Best Local Similarity 63.6%; Pred. No. 31;
 Matches 7; Conservative 2; Mismatches 0; Gaps 0;

QY 1 AGLCPDPRPL 11

DB 120 AGLVCDAPRPV 130

RESULT 7

QJ02206

U146h protein - Marek's disease virus

C:Species: Marek's disease virus

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999

C:Accession: JQ2206

R:Yanagida, N.; Yoshida, S.; Nazerian, K.; Lee, L.F.

J. Gen. Virol. 74, 1837-1845, 1993

A:Title: Nucleotide and predicted amino acid sequences of Marek's disease virus homologs

A:Reference number: JQ2199; MUID:93389438

A:Accession: JQ2206

A:Molecule type: DNA

A:Residues: 1-568 <YAN>

A:Cross-references: GB:L10283; NID:g388703; PIDN:AAA03151.1; PID:g388709

A:Experimental source: strain GA

Query Match 60.3%; Score 41; DB 2; Length 568;
 Best Local Similarity 54.5%; Pred. No. 40;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLUCPDPRPLE 12

DB 479 GVLCPAPQPAD 489

RESULT 8

S01939

hypothetical protein 2 - beet western yellows virus (isolate FL1)

C:Species: beet western yellows virus

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 20-Sep-1999

C:Accession: S01939

R:Veidt, I.; Lot, H.; Leiser, M.; Scheidecker, D.; Guillely, H.; Richards, K.; Jonard, G.

Nucleic Acids Res. 16, 9917-9932, 1988

A:Title: Nucleotide sequence of beet western yellows virus RNA.

A:Reference number: S01935; MUID:89057523

A:Accession: S01939

A:Molecule type: genomic RNA

A:Residues: 1-607 <VEI>

A:Cross-references: EMBL:X13063; NID:g62294; PIDN:CAA31463.1; PID:g62296

C:Superfamily: potato leaf roll virus 70K protein

Query Match 60.3%; Score 41; DB 2; Length 607;
 Best Local Similarity 85.7%; Pred. No. 43;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 CPDPRPL 11

DB 62 CPDPRPL 68

RESULT 9

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N:Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein e
 C:Species: Homo sapiens (man)
 C:Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
 C:Accession: A24571; A25491; A44188; B44188; I59509; I57622
 R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T
 Nature 319, 230-234, 1986
 A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
 A:Reference number: A24571; MUID:86118663

A:Accession: A24571

A:Molecule type: mRNA

A:Residues: 1-1255 <YAM>

A:Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid

A:Reference number: A25491; MUID:86016729

A:Accession: A25491

A:Molecule type: DNA

A:Residues: 737-1031 <SEM>

A:Cross-references: GB:M11767; NID:gl82163; PIDN:AAA35808.1; PID:g553282

R:Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg

Science 230, 1132-1139, 1985

A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro

A:Reference number: A44188; MUID:86070181

A:Accession: A44188

A:Molecule type: DNA

A:Residues: 740-910 <COU1>

A:Cross-references: GB:M12036; NID:gl83988; PIDN:AAA35978.1; PID:g183989

A:Accession: B44188

A:Molecule type: mRNA

A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A:Cross-references: GB:M11730; NID:gl83986

R:King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A:Reference number: I59509; MUID:85272597

A:Accession: I59509

A:Status: translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 832-909 <REX>

A:Cross-references: GB:L129395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio

A:Reference number: I57622; MUID:87286898

A:Accession: I57622

A:Status: translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-191 <TAL>

A:Cross-references: GB:M16792; NID:gl83983; PIDN:AAA58637.1; PID:g553332

C:Comment: Amplification and overexpression of this erbB-related gene occurs in about

C:Genetics:

A:Gene: GDB:ERBB2; NGL: NEU; HER-2

A:Cross-references: GDB:120613; OMIM:164870

A:Map position: 17q21.1-17q21.1

A:Introns: 25/1; 75/3; 147/1; 883/3

A>Note: the list of introns is incomplete

C:Function:

A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos

inase

F:1-21/Domain: signal sequence #status predicted <SIG>

F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>

F:22-653/Domain: extracellular #status predicted <EXT>

F:70-304/Domain: EGF receptor extracellular domain repeat <EEL>

F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>

F:654-675/Domain: transmembrane #status predicted <TMW>

F:676-1255/Domain: intracellular #status predicted <INT>

F:718-983/Domain: protein kinase homology <KIN>

F:726-734/Region: protein kinase ATP-binding motif

F:68-124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status pred

F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status pred

F:753/Active site: Lys #status predicted
F:1159,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 60.3%; Score 41; DB 1; Length 1255;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDPP 10
DB 1029 GFCCPDPP 1037

RESULT 10
G83562
C:Species: Pseudomonas aeruginosa (strain PA01)
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83562
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizioch, S.D.; Warriner, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: G83562
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-140 <STO>
A:Cross-references: GB:AE004501; GB:AE004053; PIDN:AA04053.1; GSPDB:GN001
A:Experimental source: Strain PA01
C:Genetics:
A:Gene: PA0664

Query Match 58.8%; Score 40; DB 2; Length 140;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 9; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 2 GLLCPDPP 12
DB 117 GRLCPMTASPRPL 131

RESULT 11
G84499
hypothetical protein At2g11920 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
R:Accession: G84499
M.: Koo, H.; Moffat, K.S.; Rounslev, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: G84499
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:Cross-references: GB:AE002093; NID:94557064; PIDN:AA022503.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g11920
A:Map position: 2

Query Match 58.8%; Score 40; DB 2; Length 210;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 LKCPDPP 11
DB 169 LVCPDPRS 177

RESULT 12

D70871
hypothetical protein RV1457c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: D70871
R:Coner, R.; Davies, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
K.; Cole, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: D70871
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-261 <COL>
A:Cross-references: GB:AL021184; GB:AL123456; NID:93261498; PIDN:CAA15987.1; PID:9279
A:Experimental source: strain H37RV
C:Genetics:
C:Superfamily: Streptomyces peucetius daunorubicin resistance protein

Query Match 58.8%; Score 40; DB 2; Length 261;
Best Local Similarity 70.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGULCPDPP 10
DB 11 AGTFSPPDPP 20

RESULT 13

D7301
VirB8 homolog - Bordetella pertussis
C:Species: Bordetella pertussis
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994
C:Accession: G47301
R:Weiss, A.A.; Johnson, F.D.; Burns, D.L.
A:Title: Molecular characterization of an operon required for pertussis toxin secreti
A:Reference number: A47301; MUID:93219406
A:Contents: Bp338
A:Accession: G47301
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-261
A:Notes: Sequence extracted from NCBI backbone (NCBIN:128775, NCBIP:128779)

Query Match 58.8%; Score 40; DB 2; Length 276;
Best Local Similarity 63.6%; Pred. No. 29;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGULCPDPP 11
DB 40 ARIVMPDPP 50

RESULT 14

D85942
Dehydratase ygeX [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: D85942
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
illier, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nardone, S.; et al.
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551

Qy 2 GLLCPRPL 11

J. Biol. Chem. 268, 23376-23381, 1993

A: Title: Sequence and regulation of the uapA gene encoding a uric acid-xanthine permease from *Ureaplasma urealyticum* strain A48878; PMID:94043131

A: Accession: A48878

A: Molecule type: DNA

A: Residues: 1-595 <GOR>

A: Cross-references: EMBL:X71807; NID:g298063; PID:g298064

C: Geneticks:

A: Gene: uapA

A: Introns: 61/2: 343/3: 554/2

A: Translations: uric acid/xanthine transport protein

C: Keywords: transmembrane protein

Query Match 58.1%; Score 39.5; DB 2; Length 595;

Best Local Similarity 61.5%; Pred. NO. 74;

Matches 8; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

QY 2 GLICPD--PRPL 11
1111111111

DB 273 GMLCPSATAPRPL 285

RESULT 22

A: Title: Submaxillary protein SMRL precursor - black rat

C: Species: Rattus rattus (black rat, roof rat)

A: Note: This sequence is derived from an apparently genuine specimen of this problema

A: C: Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Jun-2000

C: Accession: F53030

R: Singer, M.; Courtney, Y.; Rougeon, F.

Dev. Biol. 147:1144, 1995

A: Reference number: F53030; PMID:95169272

A: Accession: F53030

A: Status: preliminary; translated from GB/EMBL/DDBJ

A: Molecule type: mRNA

A: Residues: 1-146 <RES>

A: Cross-references: EMBL:X77919; NID:g732921; PIDN:CAA54834.1; PID:g732922

A: Geneticks: alpha

C: Superfamily: proline-rich peptide P-B

C: Keywords: glycoprotein

F: 1-18/Domain: signal sequence #status predicted <SIG>

F: 19-146/Product: submaxillary protein SMRL #status predicted <MAT>

F: 129,136/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 57.4%; Score 39; DB 2; Length 146;

Best Local Similarity 77.8%; Pred. NO. 23;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LCPDPRPL 11
1111111111

DB 98 LTAPDPRPL 106

RESULT 23

I48343

A: Title: Interleukin-11 receptor alpha-chain precursor - mouse

C: Species: Mus musculus (house mouse)

A: Note: This sequence is derived from an apparently genuine specimen of this problema

C: Accession: I48343; S51619

R: Neuhaus, H.; Bettenhausen, B.; Billinski, P.; Simon-Chazottes, D.; Guenet, J.L.; Gos

Dev. Biol. 166, S21-542, 1994

A: Title: etl2, a novel putative type-1 cytokine receptor expressed during mouse embryo

A: Reference number: I48343

A: Accession: I48343

A: Status: preliminary; translated from GB/EMBL/DDBJ

A: Residues: 1-432 <RES>

A: Cross-references: EMBL:X74953; NID:g673437; PIDN:CAA52908.1; PID:g673438

R:Hilton, D.J.; Hilton, A.A.; Raicevic, A.; Rakar, S.; Harrison-Smith, M.; Gough, N.M.;
EMBO J. 13, 4765-4775, 1994
A:Title: Cloning of a murine IL-11 receptor alpha-chain; requirement for gp130 for high
A:Reference number: S51619; MUID:95045367
A:Accession: S51619

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-432 <HIL>
A:Cross-references: EMBL:U14412; NID:g576454; PIDN:AAA53248.1; PID:g576455
C:Genetics:
A:Gene: Et12/IL11

C:Superfamily: ciliary neurotrophic factor receptor; cytokine receptor homology; immunog
C:Keywords: cytokine receptor
F:120-310/Domain: cytokine receptor homology <CR>

Query Match 57.4%; Score 39; DB 2; Length 432;
Best Local Similarity 77.8%; Pred. No. 65;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPLE 12
| | | | | | | |
Db 352 LQDPRPLD 360

RESULT 24

T04329

Importin alpha - tomato

N:Alternate names: KAP alpha protein

C:Species: Lycopersicon esculentum (tomato)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 29-Sep-1999

C:Accession: T04329

R:Kunik, T.; Mizrachy, L.; Citovsky, V.; Gafni, Y.

Submitted to the EMBL Data Library, June 1998

A:Reference number: Z15286

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Accession: T04329

A:Molecule type: mRNA

A:Residues: 1-527 <KUN>

A:Cross-references: EMBL:AF017252; NID:g3228369; PIDN:AAC23722.1; PID:g3228370

C:Genetics:

A:Gene: KAP

C:Superfamily: pendulin

Query Match 57.4%; Score 39; DB 2; Length 527;
Best Local Similarity 85.7%; Pred. No. 79;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9
| | | | | | | |
Db 422 LVCPDPR 428

RESULT 25

T52102

probable nuclear transport factor importin alpha-like protein [imported] - Arabidopsis t

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000

C:Accession: T52102

R:Merkle, T.

Submitted to the EMBL Data Library, March 1998

A:Reference number: Z25955

A:Accession: T52102

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-532 <MER>

A:Cross-references: EMBL:Y15224; PIDN:CAA75513.1

A:Experimental source: ecotype Columbia; vegetative; 3 weeks old

C:Genetics:

A:Gene: Impa1

C:Superfamily: pendulin

Query Match 57.4%; Score 39; DB 2; Length 532;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9
| | | | | | | |
Db 421 LVCPDPR 427

RESULT 26

T52268

Importin alpha [validated] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000

C:Accession: T52268

R:Smith, H.M.; Hicks, G.R.; Raikhel, N.V.

Plant Physiol. 114, 411-417, 1997

A:Title: Importin alpha from Arabidopsis thaliana is a nuclear import receptor that r

A:Reference number: Z26010; MUID:97336305

A:Accession: T52268

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-532 <SMI>

A:Cross-references: EMBL:AF077528; PIDN:AAC27644.1

A:Experimental source: Cultivar Columbia

C:Function:

A:Description: involved in protein import in nucleus; binds to nuclear localization s

C:Superfamily: pendulin

Query Match 57.4%; Score 39; DB 2; Length 532;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9
| | | | | | | |
Db 421 LVCPDPR 427

RESULT 27

T52098

probable nuclear transport factor importin alpha [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000

C:Accession: T52098

R:Schledz, M.; Leclerc, D.; Neuhaus, G.; Merkle, T.

Submitted to the EMBL Data Library, August 1997

A:Reference number: Z25951

A:Accession: T52098

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-535 <SCH>

A:Cross-references: EMBL:Y14615; PIDN:CAA74965.1

A:Experimental source: ecotype Columbia; vegetative tissue; 3 weeks old

C:Genetics:

A:Gene: Impa-2

C:Superfamily: pendulin

Query Match 57.4%; Score 39; DB 2; Length 535;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9
| | | | | | | |
Db 426 LVCPDPR 432

RESULT 28

EGHD

epidermal growth factor precursor [validated] - human

N:Alternate names: urogastrone precursor

A:Cross-references: GB:AE001800; GB:AE000512; NID:g4982090; PIDN:AA036598.1; PID:g498209
A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1531

C:Superfamily: electron transfer flavoprotein alpha chain

Query Match 55.9%; Score 38; DB 2; Length 338;

Best Local Similarity 60.0%; Pred. No. 75;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLCPDPRP 10

|||||

Db 149 ATMCPDHP 158

RESULT 31

G75571

MutR/nudix family protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

C:Accession: G75571

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: G75571

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-350 <WHI>

A:Cross-references: GB:AE001864; GB:AE000513; NID:g6457659; PIDN:AAF09597.1; PID:g645766

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0004

A:Map position: 1

Query Match 55.9%; Score 38; DB 2; Length 350;

Best Local Similarity 70.0%; Pred. No. 77;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLCPDPRPL 11

|||||

Db 72 GLCPNLRPL 81

RESULT 32

S74343

probable aspartate transaminase (EC 2.6.1.1) aspc sll0402 [similarity] - Synecocystis s

N:Alternate names: hypothetical protein sll0402

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S74343

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S74343

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-389 <KAN>

A:Cross-references: EMBL:D64001; GB:AB001339; NID:g1001102; PIDN:BAAL0261.1; PID:g100112

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: aspc

C:Superfamily: aspartate transaminase

C:Keywords: aminotransferase; phosphoprotein; pyridoxal phosphate

F:238/Binding site: pyridoxal phosphate (lys) (covalent) #status predicted

Query Match 55.9%; Score 38; DB 2; Length 389;

Best Local Similarity 66.7%; Pred. No. 86;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGLCPDPRP 9

|||||

Db 312 AGLSCPDPK 320

RESULT 33

T33943

hypothetical protein C01B4.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33943

R:Smith, A.; Wamsley, P.; Fronick, W.

submitted to the EMBL Data Library, February 1999

A:Description: The sequence of C. elegans cosmid C01B4.

A:Reference number: Z21443

A:Accession: T33943

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-475 <SMI>

A:Cross-references: EMBL:AF125952; PIDN:AA014699.1; GSPDB:GN00023; CESP:C01B4.7

A:Experimental source: strain Bristol N2; clone C01B4

C:Genetics:

A:Gene: CESP:C01B4.7

A:Map position: 5

A:Introns: 45/2; 80/1; 118/2; 189/3; 239/2; 340/3; 433/3

Query Match 55.9%; Score 38; DB 2; Length 475;

Best Local Similarity 66.7%; Pred. No. 1e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLCPDPRP 10

|||||

Db 423 GLCPDNTTP 431

RESULT 34

S74637

lysine decarboxylase cad - Synecocystis sp. (strain PCC 6803)

N:Alternate names: hypothetical protein sll1683

C:Species: Synecocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S74637

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.;

Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocys

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S74637

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-483 <KAN>

A:Cross-references: EMBL:D90900; GB:AB001339; NID:g151768; PIDN:BAAL6789.1; PID:g165

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: cad

C:Superfamily: lysine decarboxylase cad

Query Match 55.9%; Score 38; DB 2; Length 483;

Best Local Similarity 87.5%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGLCPDPP 8

|||||

Db 431 AGLICPYP 438

RESULT 35

E70645

Hypothetical protein RV3132c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: E70645

R:Colles, S. T.; Broach, R.; Parkhill, J.; Garnier, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M. A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J. E.; Taylor, K.; Whitehead, S.; Barrell, B. G.

A:Note: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference: E70645

A:Accession: E70645

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-578 <COL>

A:Cross-references: GB:Z83867; GB:AL123456; NID:G3261695; PIDN:CAB06282.1; PID:e290953;

C:Experimental source: strain H37RV

A:Gene: RV3132c

Query Match 55.9%; Score 38; DB 2; Length 578;

Best Local Similarity 70.0%; Pred. NO. 1.3e+02; Indels 0; Gaps 0;

Matches 7; Conservative 1; Mismatches 2;

OY 2 GLLCPDPRPL 11

Db 122 GLLIEDPRPL 131

RESULT 36

S27756

receptor-like protein kinase 5 (pc 2.7.1.-) precursor - Arabidopsis thaliana

N:Alternate names: protein F2009.180

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jun-1999

C:Accession: S27756; T04620

R:Submitted to the EMBL Data Library, February 1992

A:Description: Receptor-like protein kinase genes of Arabidopsis thaliana.

A:Reference number: S27754

A:Accession: S27756

A:Molecule type: mRNA

A:Residues: 1-999 <WAL>

A:Cross-references: EMBL:M84650; NID:G166849; PIDN:AAA32859.1; PID:G166850

R:Evans, M.; Rose, M.; Hempel, S.; Entian, K. D.; Hoheisel, J.; Mewes, H. W.; Mayer, K. F. X

submitted to the Protein Sequence Database, October 1998

A:Reference number: Z15380

A:Accession: T04620

A:Molecule type: DNA

A:Residues: 1-999 <BEV>

A:Experimental source: EMBL:AL021749

C:Genetics:

A:Gene: RLK5

A:Map position: 4

A:Introns: 868/1

C:Note: #2009.180

C:Keywords: ATP; autophosphorylation; glycoprotein; protein kinase; phosphotransferase; receptor

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-999/Product: receptor-like protein kinase #status predicted <MAT>

F:66-89/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>

F:90-114/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>

F:115-139/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>

F:140-167/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>

F:168-187/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>

F:188-212/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>

F:213-260/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>

F:237-260/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>

F:261-284/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>

F:285-307/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR10>

F:308-331/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR11>

F:332-355/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR12>

F:356-379/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR13>

F:380-403/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR14>

F:404-427/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR15>

F:428-475/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR16>

F:476-499/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR17>

F:500-523/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR18>

F:524-547/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR19>

F:548-571/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR20>

F:571-593/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR21>

F:681-971/Region: protein kinase ATP-binding motif

F:689-697/Region: protein kinase ATP-binding motif

F:981,102,150,185,210,269,282,452,576/Binding site: carbohydrate (Asn) (covalent) #sta

F:711,737,819,821/Active site: Lys, Glu, Asp, Lys #status predicted

F:824,828/Binding site: magnesium (Asn, Asp) #status predicted

Query Match 55.9%; Score 38; DB 1; Length 999;

Best Local Similarity 70.0%; Pred. NO. 2.1e+02; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 3;

OY 2 GLLCPDPRPL 11

Db 943 GLLCTSPPL 952

RESULT 37

S64146

Nucleic acid membrane protein YGL133w - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein G2842

C:Species: Saccharomyces cerevisiae

C:Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 17-Apr-1998

C:Accession: S64146; S71739

R:Escaribano, V.; Eraso, P.; Portillo, F.; Mazon, M. J.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64146

A:Accession: S64146

A:Molecule type: DNA

A:Residues: 1-1264 <ESC>

A:Cross-references: EMBL:Z72655; NID:G1322701; PID:e243361; PID:G1322702; MIPS:YGL133

A:Experimental source: strain S288C

R:Escaribano, V.; Eraso, P.; Portillo, F.; Mazon, M. J.

Yeast 12, 887-892, 1996

A:Fitte: Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces cerevisiae chro

A:Reference number: S71733; MUID:96437978

A:Accession: S71739

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1264 <ESW>

A:Cross-references: EMBL:X92670

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A:Map position: 7L

C:Keywords: transmembrane protein

F:549-565/Domain: transmembrane #status predicted <TMM>

Query Match: 55.9%; Score 38; DB 2; Length 1264;

Best Local Similarity 66.7%; Pred. NO. 2.7e+02; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 1;

OY 3 LLCDDPRPL 11

Db 9 ILCPDPRPL 17

RESULT 38

JEO352
mucin MUC5B, tracheobronchial - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C:Accession: JEO352
R:Offner, G.D.; Nunes, D.P.; Keates, A.C.; Afghal, N.H.; Troxler, R.F.
Biochem. Biophys. Res. Commun. 251, 350-355, 1998
A:Title: The amino-terminal sequence of MUC5B contains conserved multifunctional D domain
A:Reference number: JEO352; MUID:99009274
A:Accession: JEO352
A:Molecule type: mRNA
A:Residues: 1-1321 <OFF>
A:CROSS-references: GB:AF086604; NID:g3789926; PIDN:AAC67545.1; PID:g3789927
C:Comment: This protein is large multimeric glycoproteins which is secreted by epithelial cells
C:Genetics:
A:Gene: MUC5B

Query Match 55.9%; Score 38; DB 2; Length 1321;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 CPDPRPL 11

|||||

Db 244 CPDPLPL 250

RESULT 39

TJ35694
ATP dependent DNA helicase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: TJ35694
R:Harris, D.; Taylor, K.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1998
A:Reference number: 221587
A:Accession: TJ35694
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1690 <HAR>
A:CROSS-references: EMBL:AL031031; PIDN:CAA19862.1; GSPDB:GN00070; SCOEDB:SC7C7.16C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC7C7.16C

Query Match 55.9%; Score 38; DB 2; Length 1690;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLCPDPRPLE 12

|||||

Db 1204 APLLPPHPHLE 1215

RESULT 40

Ti8296
myosin heavy chain - Entamoeba histolytica
C:Species: Entamoeba histolytica
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C:Accession: Ti8296
R:Guillen, N.
submitted to the EMBL Data Library, February 1997
A:Reference number: 218855
A:Accession: Ti8296
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2139 <GUI>
A:CROSS-references: EMBL:L03534; NID:g1850912; PID:g1850913; PIDN:AAB48065.1
C:Genetics:
A:Gene: mhca

C:Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

Query Match 55.9%; Score 38; DB 2; Length 2139;
Best Local Similarity 66.7%; Pred. No. 4.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10

|:| |:| |

Db 24 GILIPNRP 32

RESULT 41

B83432
conserved hypothetical protein in type III secretion PA1701 [imported] - Pseudomonas
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83432
R:Stover, C.K.; Yuan, Y.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: B83432
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-121 <STO>
A:CROSS-references: GB:AE004597; GB:AE004091; NID:g9947671; PIDN:AAG05090.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1701
C:Superfamily: Yersinia pestis hypothetical protein Y0045

Query Match 54.4%; Score 37; DB 2; Length 121;
Best Local Similarity 70.0%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPDPRPLE 12

|:| |:| |

Db 31 LLAPDGRPVE 40

RESULT 42

S63639
hypothetical protein 246 - Allomyces macrogynus mitochondrion
C:Species: mitochondrion Allomyces macrogynus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
C:Accession: S63639
R:Paquin, B.; Lang, B.F.
J. Mol. Biol. 255, 688-701, 1996
A:Title: The mitochondrial DNA of Allomyces macrogynus: the complete genomic sequence
A:Reference number: S63635; MUID:96226032
A:Accession: S63639
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-246 <PAQ>
A:CROSS-references: EMBL:U41288
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995.
C:Genetics:
A:Genome: mitochondrion
A:Introns: 243/3
C:Superfamily: probable RNA maturase
C:Keywords: mitochondrion

Query Match 54.4%; Score 37; DB 2; Length 246;
Best Local Similarity 75.0%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPL 11

Db 99 LAPDRPL 106
RESULT 43
H85408
GATA transcription factor 3 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Accession: H85408
C:Date: 23-Apr-1999 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Accession: H85408
A:Residues: 1-269 <STC>
A:Molecule type: DNA
A:Cross-references: GB:NC_001268; NID:97270419; PIDN:CA80185.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g34680
A:Map position: 4
C:Superfamily: Arabidopsis thaliana GATA transcription factor 4
Query Match 54.4%; Score 37; DB 2; Length 269;
Best Local Similarity 50.0%; Pred. No. 87;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
OY 1 AGLCPDRPLE 12
DB 213 SGRCLPEVRPAD 224
RESULT 44
T05288
GATA-binding transcription factor homolog 3 [imported] - Arabidopsis thaliana
N:Alternate names: protein T4L20.260
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 31-Dec-2000
C:Accession: T05288; T52105
R:Bevan, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.; De
submitted to the protein Sequence Database, September 1998
A:Reference number: Z15406
A:Accession: T05288
A:Molecule type: DNA
A:Residues: 1-269 <BEV>
A:Cross-references: EMBL:AL023094
A:Genetics: B1
R:Teakle, G.R.; Gilmartin, P.M.
submitted to the EMBL Data Library, June 1997
A:Description: Two types of GATA factor are found in fungi but are uniquely represented
A:Reference number: Z25956
A:Accession: T52105
A:Molecule type: mRNA
A:Cross-references: translated from GB/EMBL/DDBJ
A:Residues: 1-54, 'Q', 56-269 <TEA>
A:Cross-references: EMBL:X13650; PIDN:CAA74001.1
A:Experimental source: ecotype Columbia
C:Genetics: B2
A:Gene: T4L20.260
A:Map position: 4
A:Introns: 94/3
A:Note: T4L20.260
C:Genetics: <B2>
A:Gene: GATA-3
C:Superfamily: Arabidopsis thaliana GATA transcription factor 4
Query Match 54.4%; Score 37; DB 2; Length 269;

Best Local Similarity 50.0%; Pred. No. 87;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
OY 1 AGLCPDRPLE 12
DB 213 SGRCLPEVRPAD 224
RESULT 45
S28088
gene B protein - yeast (Kluyveromyces marxianus var. drosophilum) plasmid pKD1
C:Species: Kluyveromyces marxianus var. drosophilum
C:Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 29-Oct-1999
C:Accession: S28088
R:Chen, X.J.; Salioia, M.; Falcone, G.; Bianchi, M.M.; Fukuhara, H.
A:Title: Sequence organization of the circular plasmid pKD1 from the yeast Kluyveromy
A:Reference number: S28086; MUID:86232585
A:Accession: S28088
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-415 <CHS>
A:Cross-references: EMBL:X03961; NID:92800; PIDN:CAA27593.1; PID:92803
C:Genetics:
A:Gene: B
A:Genome: plasmid
Query Match 54.4%; Score 37; DB 2; Length 415;
Best Local Similarity 70.0%; Pred. No. 13e+02;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 3 LLCPPDRPLE 12
DB 132 LLCPSPELLE 141
Search completed: June 28, 2001, 11:56:59
Job time: 240 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:22 ; Search time 22.61 seconds
(without alignments)
18.181 Million cell updates/sec

Title: US-09-439-313-558
Perfect score: 68
Sequence: 1 AGLCPDPRLE 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	67.6	567	1 NARQ_HAEIN	P44604 haemophilus
2	44	64.7	397	1 YC64_MYCTU	Q11055 mycobacteri
3	41	60.3	607	1 V66K_BWYVF	P09506 beet wester
4	41	60.3	1255	1 ERB2_HUMAN	P04626 homo sapien
5	40	58.8	398	1 DPAL_ECOLI	Q46804 escherichia
6	39.5	58.1	615	1 UAPA_EMENI	Q07307 emericella
7	39	57.4	527	1 IMA_LYCES	O22478 lycopersico
8	39	57.4	596	1 IMA1_ARATH	O96321 arabidopsis
9	39	57.4	1207	1 EGF_HUMAN	P01133 homo sapien
10	39	57.4	1687	1 Z142_HUMAN	P52746 homo sapien
11	38	55.9	338	1 ETFA_MEGEL	O85692 megasphaera
12	38	55.9	389	1 AAT_SYNY3	Q55128 synechocyst
13	38	55.9	395	1 DIHR_MANSE	P35464 manduca sex
14	38	55.9	999	1 RLK5_ARATH	P47735 arabidopsis
15	38	55.9	1264	1 YGN3_YEAST	P53125 saccharomyc
16	37	54.4	391	1 P53_CAVPO	Q9WU6 cavia porce
17	37	54.4	415	1 REP1_KLULA	P13775 kluyveromyc
18	37	54.4	518	1 BAR2_SCHCO	Q05659 schizophyll
19	37	54.4	567	1 GPV_MOUSE	O08742 mus musculu
20	37	54.4	653	1 LEPA_MYCTU	P71739 mycobacteri
21	37	54.4	659	1 NRTC_SYNP7	P38045 synechococc
22	36	52.9	229	1 UNG_CHLTR	O84613 chlamydia t
23	36	52.9	291	1 IBP3_BOVIN	P20959 bos taurus
24	36	52.9	352	1 TCBF_PSESO	P27101 pseudomonas
25	36	52.9	352	1 TFEI_BURCE	Q45072 burkholderi
26	36	52.9	354	1 TFF1_ALCEU	P27137 alcaligenes
27	36	52.9	356	1 HX2_HUMAN	P14652 homo sapien
28	36	52.9	1022	1 CA26_CHICK	P15988 gallus gall
29	36	52.9	1212	1 UTY_MOUSE	P79457 mus musculu
30	36	52.9	1337	1 YDM5_SCHPO	P87136 schizosacch
31	36	52.9	1572	1 BA12_HUMAN	O60241 homo sapien
32	36	52.9	2124	1 Y192_HUMAN	Q93074 homo sapien
33	35.5	52.2	249	1 RU2A_ARATH	P43333 arabidopsis

34	35.5	52.2	501	1 GSPE_AERHY	P31741 aeromonas h
35	35	51.5	85	1 PLF4_SHEEP	P30035 ovis aries
36	35	51.5	136	1 SR19_ORYSA	P49964 oryza sativ
37	35	51.5	266	1 IBP3_PIG	P16611 sus scrofa
38	35	51.5	267	1 CYSH_PSEAE	O05927 pseudomonas
39	35	51.5	271	1 YPOQ_KLEPN	P75099 klebsiella
40	35	51.5	299	1 FCN3_HUMAN	O75636 homo sapien
41	35	51.5	319	1 RM04_YEAST	P36517 saccharomyc
42	35	51.5	330	1 ETFA_CLOTS	P71153 clostridium
43	35	51.5	348	1 YT35_STRFR	P20186 streptomyce
44	35	51.5	387	1 UL33_RCMVM	O12000 rat cytomeg
45	35	51.5	454	1 GLNA_HALVO	P43386 halobacteri
46	35	51.5	481	1 MSK_MOUSE	Q60670 mus musculu
47	35	51.5	543	1 PKAA_STRCO	P54739 streptomyce
48	35	51.5	553	1 DMWD_HUMAN	O09019 homo sapien
49	35	51.5	573	1 YKK3_CAEEL	P34280 caenorhabdi
50	35	51.5	653	1 HGFA_MOUSE	Q90988 mus musculu
51	35	51.5	704	1 RP3A_BOVIN	O06846 bos taurus
52	35	51.5	714	1 HUNK_HUMAN	P57058 homo sapien
53	35	51.5	776	1 SNIL_RAT	Q91145 rattus norv
54	35	51.5	778	1 TAST_HUMAN	Q12815 homo sapien
55	35	51.5	786	1 SNIL_HUMAN	P57059 homo sapien
56	35	51.5	1019	1 LFC_TACTR	P28175 tachypleus
57	35	51.5	1029	1 CA26_MOUSE	Q02788 mus musculu
58	35	51.5	1162	1 LEPR_RAT	Q62959 rattus norv
59	35	51.5	2871	1 DESP_HUMAN	P15924 homo sapien
60	34	50.0	77	1 Y188_ARCFU	O30050 archaeoglob
61	34	50.0	121	1 SECR_HUMAN	P09683 homo sapien
62	34	50.0	162	1 PHAC_CYAPA	P16282 cyanophora
63	34	50.0	178	1 BTC_HUMAN	P35070 homo sapien
64	34	50.0	188	1 ADML_CANFA	O75559 canis famil
65	34	50.0	205	1 YK07_YEAST	P36061 saccharomyc

ALIGNMENTS

RESULT	1	NARQ_HAEIN	STANDARD;	PRT;	567 AA.
ID	NARQ_HAEIN				
AC	P44604;				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	SENSOR PROTEIN NARQ HOMOLOG (EC 2.7.3.-)				
GN	NARQ OR HI0267.				
OS	Haemophilus influenzae.				
OC	Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;				
OC	Haemophilus.				
OX	NCBI_TaxID=727;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=RD / KW20 / ATCC 51907;				
RA	MEDLINE=95350630; PubMed=7542800;				
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,				
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,				
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,				
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,				
RA	Wiedman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,				
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,				
RA	Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,				
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,				
RA	Venter J.C.;				
RT	"Whole-genome random sequencing and assembly of Haemophilus				
RL	influenzae Rd. "				
CC	Science 269:496-512(1995).				
CC	-!- FUNCTION: PROBABLE MEMBER OF A TWO-COMPONENT REGULATORY SYSTEM;				
CC	HOWEVER IT IS NOT KNOWN WHAT PROTEIN IT PHOSPHORYLATES AND IN WHICH				
CC	REGULATORY PATHWAY IT ACTS AS THE NARL AND OTHER NAR GENES DO NOT				
CC	EXIST IN H. INFLUENZAE.				
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE				
CC	(PROBABLE).				
CC	-!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE				

CC CC KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is not modified

CC and this statement is included in the content of any copy made for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U32713; AAC21933.1; -

CC TIGR; H10267; -

CC InterPro; IPR000410; -

CC Pfam; PF00512; DUF51; 1.

CC Pfam; PF00512; signal; 1.

CC Sensory transduction; Transferase; Kinase; Phosphorylation;

CC Transmembrane; Inner membrane.

CC DOMAIN 1 13

CC TRANSMEM 1 13

CC TRANSMEM 35 147

CC TRANSMEM 148 172

CC TRANSMEM 173 567

CC DOMAIN 350 567

CC MOD_RES 373 373

CC SEQUENCE 567 AA; 65238 MW; A74F549986CECAB9 CRC64;

CC -----

CC Query Match 67.6%; Score 46; DB 1; Length 567;

CC Best Local Similarity 66.7%; Pred. NO. 2.6;

CC Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

CC

CC 1 AGLCPDPRPL 12

CC 326 AGLCPDPRIMQ 337

CC

CC RESULT 2

CC ID YC64 MYCTU STANDARD; PRT; 397 AA.

CC AC 01-OCT-1996 (Rel. 34, Created)

CC DT 01-OCT-1996 (Rel. 34, Last sequence update)

CC DE 01-OCT-2000 (Rel. 40, Last annotation update)

CC DE HYPOTHETICAL 42.2 KDA PROTEIN RV1264.

CC GN RV1264 OR MTCV50.18C.

CC OS Mycobacterium tuberculosis

CC OC Actinobacteriales; Actinobacteria; Actinobacteridae;

CC OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

CC OX NCBI_Taxid=1773;

CC [1]

CC SEQUENCE FROM N.A.

CC RC STRAIN-H37RV;

CC RX MEDLINE=98295987; PubMed-9634230;

CC RA Cole S.V., Brechtner S., Parkhas S., Barry C.E. III, Tekala F.,

CC RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

CC RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

CC RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

CC RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

CC RA Rutter S., Seeger K., Skelton S., Squares S., Sultson J.E.,

CC RA Taylor K.I., Whitehead S., Barrell B.G.; Mycobacterium tuberculosis from the

CC RA complete genome sequence.

CC RL Nature 393:537-544(1998).

CC -!- SIMILARITY: CONTAINS 1 GUANYLATE CYCLASE DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is not modified

CC and this statement is included in the content of any copy made for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; 277137; CAB00890.1; -

CC HSPB; Q02846; IAWL.

CC InterPro; IPR000410; -

CC Pfam; PF00512; DUF51; 1.

CC PROSITE; PS00125; GUANYLATE_CYCLASES_2; 1.

CC HYPOTHETICAL protein.

CC SEQUENCE 397 AA; 42232 MW; F6C212A181DB5AD3 CRC64;

CC -----

CC Query Match 64.7%; Score 44; DB 1; Length 397;

CC Best Local Similarity 77.8%; Pred. NO. 4;

CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

CC

CC 3 LCPDPRPL 11

CC 269 LCPDPRPL 277

CC

CC RESULT 3

CC ID V66K BWYVF STANDARD; PRT; 607 AA.

CC AC P05506; 1989 (Rel. 10, Created)

CC DT 01-MAR-1989 (Rel. 10, Last sequence update)

CC DE 01-OCT-1989 (Rel. 12, Last annotation update)

CC DE 66.2 KDA PROTEIN (ORF 2).

CC OS Beet western yellows virus (isolate FL-1) (BWYV).

CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;

CC OC Poliovirus; 12043;

CC OX NCBI_Taxid=12043;

CC [1]

CC SEQUENCE FROM N.A.

CC RX MEDLINE=89057523; PubMed-3194229;

CC RA Veidt I., Lot H., Leiser M., Scheidecker D., Guillely H., Richards K.,

CC RA Jonard C.;

CC RL "Nucleotide sequence of beet western yellows virus RNA."

CC RL Nucleotide sequence of beet western yellows virus RNA.

CC RL SUBMITTER TO POPATO LEAFROLL VIRUS ORF2.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is not modified

CC and this statement is included in the content of any copy made for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X13063; CAA31463.1; -

CC PIR; S01939; S01939.

CC InterPro; IPR000382;

CC Pfam; PF00512; DUF51; 1.

CC PROSITE; PS00125; GUANYLATE_CYCLASES_2; 1.

CC SEQUENCE 607 AA; 66210 MW; 2A41F82911DEC84F CRC64;

CC -----

CC Query Match 60.3%; Score 41; DB 1; Length 607;

CC Best Local Similarity 85.7%; Pred. NO. 18;

CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CC

CC 5 CPDPRPL 11

CC 62 CPEPRPL 68

CC

CC RESULT 4

CC ID FBZ_HUMAN STANDARD; PRT; 1255 AA.

CC AC P04626;

CC DT 13-AUG-1987 (Rel. 05, Created)

CC DT 13-AUG-1987 (Rel. 05, Last sequence update)

CC DT 01-OCT-2000 (Rel. 40, Last annotation update)

CC

CC CC or send an email to license@isb-sib.ch).

CC EMBL; 277137; CAB00890.1; -

CC HSPB; Q02846; IAWL.

CC InterPro; IPR000410; -

CC Pfam; PF00512; DUF51; 1.

CC PROSITE; PS00125; GUANYLATE_CYCLASES_2; 1.

CC HYPOTHETICAL protein.

CC SEQUENCE 397 AA; 42232 MW; F6C212A181DB5AD3 CRC64;

CC -----

CC Query Match 64.7%; Score 44; DB 1; Length 397;

CC Best Local Similarity 77.8%; Pred. NO. 4;

CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

CC

CC 3 LCPDPRPL 11

CC 269 LCPDPRPL 277

CC

CC RESULT 3

CC ID V66K BWYVF STANDARD; PRT; 607 AA.

CC AC P05506; 1989 (Rel. 10, Created)

CC DT 01-MAR-1989 (Rel. 10, Last sequence update)

CC DE 01-OCT-1989 (Rel. 12, Last annotation update)

CC DE 66.2 KDA PROTEIN (ORF 2).

CC OS Beet western yellows virus (isolate FL-1) (BWYV).

CC OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;

CC OC Poliovirus; 12043;

CC OX NCBI_Taxid=12043;

CC [1]

CC SEQUENCE FROM N.A.

CC RX MEDLINE=89057523; PubMed-3194229;

CC RA Veidt I., Lot H., Leiser M., Scheidecker D., Guillely H., Richards K.,

CC RA Jonard C.;

CC RL "Nucleotide sequence of beet western yellows virus RNA."

CC RL Nucleotide sequence of beet western yellows virus RNA.

CC RL SUBMITTER TO POPATO LEAFROLL VIRUS ORF2.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is not modified

CC and this statement is included in the content of any copy made for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X13063; CAA31463.1; -

CC PIR; S01939; S01939.

CC InterPro; IPR000382;

CC Pfam; PF00512; DUF51; 1.

CC PROSITE; PS00125; GUANYLATE_CYCLASES_2; 1.

CC SEQUENCE 607 AA; 66210 MW; 2A41F82911DEC84F CRC64;

CC -----

CC Query Match 60.3%; Score 41; DB 1; Length 607;

CC Best Local Similarity 85.7%; Pred. NO. 18;

CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

CC

CC 5 CPDPRPL 11

CC 62 CPEPRPL 68

CC

CC RESULT 4

CC ID FBZ_HUMAN STANDARD; PRT; 1255 AA.

CC AC P04626;

CC DT 13-AUG-1987 (Rel. 05, Created)

CC DT 13-AUG-1987 (Rel. 05, Last sequence update)

CC DT 01-OCT-2000 (Rel. 40, Last annotation update)

CC

DE RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2 PRECURSOR (EC 2.7.1.112)
DE (P185RBB2) (NEU PROTO-ONCOGENE) (C-ERBB-2) (TYROSINE KINASE-TYPE CELL
DE SURFACE RECEPTOR HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Samba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.,
RA "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Samba K., Kanata N., Toyoshima K., Yamamoto T.;
RA "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RA "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -!- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-655) HAS A FREQUENCY
CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M11767; AAA35808.1; -;
DR EMBL; M11761; AAA35808.1; JOINED.
DR EMBL; M11762; AAA35808.1; JOINED.
DR EMBL; M11763; AAA35808.1; JOINED.
DR EMBL; M11764; AAA35808.1; JOINED.
DR EMBL; M11765; AAA35808.1; JOINED.
DR EMBL; M11766; AAA35808.1; JOINED.
DR EMBL; M11730; AAA75493.1; -;

DR EMBL; M12036; AAA35978.1; -;
DR EMBL; X03363; CAA27060.1; -;
DR PIR; A25491; A25491.
DR PIR; A24571; A24571.
DR HSSP; P11362; IFGI.
DR MIM; 164870; -;
DR InterPro; IPR000494; -;
DR InterPro; IPR000719; -;
DR InterPro; IPR001245; -;
DR InterPro; IPR002174; -;
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_Ldomain; 2.
DR Pfam; PF00069; pkinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Polymorphism.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 653 675 POTENTIAL.
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 654 654 I -> V.
FT VARIANT 655 655 /FTID-VAR_004077.
FT VARIANT 655 655 I -> V.
FT CONFLICT 1170 1170 /FTID-VAR_004078.
FT SEQUENCE 1255 AA; 137909 MW; 39E9DFDA04DCF962 CRC64;
SQ
Query Match 60.3%; Score 41; DB 1; Length 1255;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 GLICPPRP 10
Db 1029 GFPCDPAP 1037
RESULT 5
ID DPAL_ECOLI STANDARD; PRT; 398 AA.
AC Q46804;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PUTATIVE DIAMINOPROPIONATE AMMONIA-LYASE (EC 4.3.1.15)
DE (DIAMINOPROPIONATASE) (ALPHA,BETA-DIAMINOPROPIONATE AMMONIA-LYASE).
GN YGEX.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC of the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X71807; GAA50681.1; --
 CC DR InterPro:IPR000444; --
 CC DR Pfam: PF00860; xan_ur_permease; 1.
 CC DR PROSITE: PS01116; XANTH_URACIL_TRANSPORTER; 1.
 CC KW TRANSMEM 118 138 POTENTIAL.
 CC FT TRANSMEM 152 172 POTENTIAL.
 CC FT TRANSMEM 172 192 POTENTIAL.
 CC FT TRANSMEM 226 250 POTENTIAL.
 CC FT TRANSMEM 258 278 POTENTIAL.
 CC FT TRANSMEM 305 325 POTENTIAL.
 CC FT TRANSMEM 337 356 POTENTIAL.
 CC FT TRANSMEM 379 402 POTENTIAL.
 CC FT TRANSMEM 468 488 POTENTIAL.
 CC FT TRANSMEM 512 532 POTENTIAL.
 CC FT TRANSMEM 563 583 POTENTIAL.
 CC FT CARBOHYD 10 10 N-LINKED (GLCNAC..) (POTENTIAL).
 CC SEQUENCE 615 AA; 65453 MW; E56A984D956897E7 CRC64;
 CC
 CC Query Match 58.18; Score 39.5; DB 1; Length 615;
 CC Best Local Similarity 61.5%; Pred No. 2;
 CC Matches Conservative 1; Mismatches 3; Gaps 3;
 CC
 CC QY 2 GLICPD--PRPL 11
 CC Db 293 GMLCPSATAPRPL 305
 CC 1:1111 1111
 CC
 CC RESULT 7
 CC ID INA_LYCES STANDARD; PRT; 527 AA.
 CC AC 022478;
 CC DT 15-JUL-1998 (Rel. 36, Created)
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)
 CC DT 01-OCT-2000 (Rel. 40, Last annotation update)
 CC DE IMPORTIN ALPHA SUBUNIT (KARYOPHERIN ALPHA SUBUNIT) (KAP ALPHA).
 CC EN Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
 CC OC Solanales; Solanaceae; Solanum.
 CC ON NCBI_TaxID=4081;
 CC RX [1]
 CC RP SEQUENCE FROM N.A. L. Citovsky V., Gafni Y.;
 CC RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC RL 1- FUNCTION: BINDS SPECIFICALLY AND DIRECTLY TO SUBSTRATES CONTAINING
 CC EITHER A SINGLE OR BIPARTITE NLS MOTIF. PROMOTES DOCKING OF IMPORT
 CC SUBSTRATES TO THE NUCLEAR ENVELOPE. SEEMS TO ACT AS A CYTOSOLIC
 CC RECEPTOR FOR BOTH SINGLE AND BIPARTITE NLS MOTIFS (BY SIMILARITY).
 CC -1- SUBUNIT: FORMS A COMPLEX WITH IMPORTIN BETA-1 SUBUNIT (BY
 CC SIMILARITY). LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE IMPORTIN ALPHA FAMILY.
 CC -1- SIMILARITY: CONTAINS 8 ARM REPEATS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC of the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC

```

DR EMBL; AF017252; AAC23722.1; -.
DR InterPro; IPR000225; -.
DR DR InterPro; IPR002652; -.
DR Pfam; PF00514; Armadillo_seg; 8.
DR Pfam; PF01749; IBB; 1.
DR KW PROSITE; PS0176; ARM_REPEAT; 5.
DR TRANSPOST; Protein transport; Repeat.
FT DOMAIN 12 51 IBB.
FT REPEAT 109 151 ARM 1.
FT REPEAT 152 196 ARM 2.
FT REPEAT 197 234 ARM 3.
FT REPEAT 235 279 ARM 4.
FT REPEAT 280 319 ARM 5.
FT REPEAT 320 362 ARM 6.
FT REPEAT 363 403 ARM 7.
FT REPEAT 403 445 ARM 8.
FT DOMAIN 446 527 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 527 AA; 58605 MW; 4A3F01691CEFA817 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 527;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9
DB 422 LVCPDPR 428

RESULT 8
ID IMAL ARATH STANDARD; PRT; 596 AA.
AC Q96321;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE IMPORTIN ALPHA-1 SUBUNIT (KARYOPHERIN ALPHA-1 SUBUNIT) (KAP ALPHA).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Ballas N., Citovsky V.;
RT "Atkapalpha gene from Arabidopsis encodes a protein that mediates
RT nuclear import of Agrobacterium VirD2 protein.";
RL (In) Plant Gene Register PGR97-129.
CC -!- FUNCTION: BINDS SPECIFICALLY AND DIRECTLY TO SUBSTRATES CONTAINING
CC EITHER A SIMPLE OR BIPARTITE NLS MOTIF. PROMOTES DOCKING OF IMPORT
CC SUBSTRATES TO THE NUCLEAR ENVELOPE. SEEMS TO ACT AS A CYTOSOLIC
CC RECEPTOR FOR BOTH SIMPLE AND BIPARTITE NLS MOTIFS (BY SIMILARITY).
CC CELLULAR RECEPTOR FOR THE NUCLEAR IMPORT OF THE VIRD2 PROTEIN OF
CC AGROBACTERIUM.
CC -!- SUBUNIT: FORMS A COMPLEX WITH IMPORTIN BETA-1 SUBUNIT (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE IMPORTIN ALPHA FAMILY.
CC -!- SIMILARITY: CONTAINS 8 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U69533; AAB72116.1; -.
DR HSSP; Q02248; 2BCT.
DR InterPro; IPR000225; -.
DR Pfam; PF00514; Armadillo_seg; 8.

```

```

DR Pfam; PF01749; IBB; 1.
DR PROSITE; PS0176; ARM_REPEAT; 3.
DR TRANSPOST; Protein transport; Repeat.
FT DOMAIN 12 51 IBB.
FT REPEAT 109 151 ARM 1.
FT REPEAT 152 196 ARM 2.
FT REPEAT 197 234 ARM 3.
FT REPEAT 235 279 ARM 4.
FT REPEAT 280 319 ARM 5.
FT REPEAT 320 362 ARM 6.
FT REPEAT 363 403 ARM 7.
FT REPEAT 403 445 ARM 8.
FT DOMAIN 446 596 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 596 AA; 65606 MW; 2A2689E1C28F43E7 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 596;
Best Local Similarity 85.7%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLCPPDR 9
DB 422 LVCPDPR 428

RESULT 9
ID EGF_HUMAN STANDARD; PRT; 1207 AA.
AC P01133;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PRO-EPIDERMAL GROWTH FACTOR PRECURSOR (EGF) [CONTAINS: EPIDERMAL
DE GROWTH FACTOR (UROGASTRONE)].
GN EGF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=87066721; PubMed=3491360;
RA Bell G.I., Fong N.M., Stempien M.M., Wormsted M.A., Caput D.,
RA Ku L., Urdea M.S., Rall L.B., Sanchez-Pescador R.;
RT "Human epidermal growth factor precursor: cDNA sequence, expression
RT in vitro and gene organization.";
RL Nucleic Acids Res. 14:8427-8446(1986).
RN [2]
RP SEQUENCE OF 971-1023.
RX MEDLINE=77117897; PubMed=300079;
RA Gregory H., Preston B.M.;
RT "The primary structure of human urogastrone.";
RL Int. J. Pept. Protein Res. 9:107-118(1977).
RN [3]
RP SEQUENCE OF 971-1023.
RX MEDLINE=89391964; PubMed=2789514;
RA Furuya M., Akashi S., Hirayama K.;
RT "The primary structure of human EGF produced by genetic engineering,
RT studied by high-performance tandem mass spectrometry.";
RL Biochem. Biophys. Res. Commun. 163:1100-1106(1989).
RN [4]
RP STRUCTURE BY NMR OF EGF.
RX MEDLINE=92395667; PubMed=1522591;
RA Hommel U., Harvey T.S., Driscoll P.C., Campbell I.D.;
RT "Human epidermal growth factor. High resolution solution structure
RT and comparison with human transforming growth factor alpha.";
RL J. Mol. Biol. 227:271-282(1992).
CC -!- FUNCTION: THE GROWTH FACTOR STIMULATES THE GROWTH OF VARIOUS
CC EPIDERMAL AND EPITHELIAL TISSUES IN VIVO AND IN VITRO AND OF SOME
CC FIBROBLASTS IN CELL CULTURE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 9 EGF-LIKE DOMAINS.

```



```
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 163 1643
FT ZN_FING 163 185
FT ZN_FING 185 242
FT ZN_FING 219 242
FT ZN_FING 253 275
FT ZN_FING 286 311
FT ZN_FING 316 340
FT ZN_FING 343 366
FT ZN_FING 372 395
FT ZN_FING 401 423
FT ZN_FING 429 451
FT ZN_FING 457 479
FT ZN_FING 485 507
FT ZN_FING 512 536
FT ZN_FING 544 567
FT ZN_FING 573 596
FT ZN_FING 1135 1158
FT ZN_FING 1171 1194
FT ZN_FING 1200 1222
FT ZN_FING 1228 1251
FT ZN_FING 1257 1280
FT ZN_FING 1286 1309
FT ZN_FING 1328 1351
FT ZN_FING 1354 1377
FT ZN_FING 1380 1403
FT ZN_FING 1424 1446
FT ZN_FING 1452 1474
FT ZN_FING 1480 1502
FT ZN_FING 1508 1530
FT ZN_FING 1536 1559
FT ZN_FING 1565 1587
FT ZN_FING 1593 1615
FT ZN_FING 1621 1643
FT CONFLICT 460 463
FT CONFLICT 541 541
FT SEQUENCE 1687 AA; 187865 MW; 42847520F04D38E7 CRC64;

Query Match 57.4%; Score 39; DB 1; Length 1687;
Best Local Similarity 70.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY - 2 GLLCPDPRP 11
DB 1334 GLLCPSPASP 1343

RESULT 11
ID ETFA_MEGEL STANDARD; PRT; 338 AA.
AC O85692;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ELECTRON TRANSFER FLAVOPROTEIN ALPHA-SUBUNIT (ALPHA-ETF) (ELECTRON
DE TRANSFER FLAVOPROTEIN LARGE SUBUNIT) (ETFLS).
GN ETFA.
OS Megasphaera elsdenii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Sporomusa subbranch;
OC Megasphaera.
OX NCBI_TaxID=907;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98361974; PubMed=9694853;
RA O'Neill H., Mayhew S.G., Butler G.;
RT Cloning and analysis of the genes for a novel electron-transferring
RT flavoprotein from Megasphaera elsdenii. Expression and
RT characterization of the recombinant protein.;
RL J. Biol. Chem. 273:21015-21024(1998).
CC -!- FUNCTION: THE ELECTRON TRANSFER FLAVOPROTEIN SERVES AS A SPECIFIC
CC ELECTRON ACCEPTOR FOR OTHER DEHYDROGENASES. IT TRANSFERS THE

ELECTRONS TO THE MAIN RESPIRATORY CHAIN VIA ETF-UBIQUINONE
OXIDOREDUCTASE (ETF DEHYDROGENASE) (BY SIMILARITY).
-!- COFACTOR: CONTAINS TWO MOLECULES OF FAD PER DIMER.
-!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
-!- SIMILARITY: BELONGS TO THE ETF ALPHA-SUBUNIT / FLXB FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL: AF072475; AAC31170.1;
DR InterPro; IPR001308;
DR Pfam; PF00766; ETF_alpha; 1.
DR PROSITE; PS00696; ETF_ALPHA; FALSE_NEG.
KW Electron transport; Flavoprotein; FAD.
FT NP_BIND 275 303 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 338 AA; 36124 MW; 63FBD4CCF111AE77 CRC64;

Query Match 55.9%; Score 38; DB 1; Length 338;
Best Local Similarity 60.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AGLLCPDPRP 10
DB 150 AETICPDNRP 159

RESULT 12
ID AAT_SYNY3 STANDARD; PRT; 389 AA.
AC Q55128;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ASPARTATE AMINOTRANSFERASE (EC 2.6.1.1) (TRANSMINASE A) (ASPAT).
GN ASPC OR SLL0403.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.;
RL DNA Res. 2:153-166(1995).
CC -!- CATALYTIC ACTIVITY: L-ASPARTATE + 2-OXOGLUTARATE = OXALOACETATE +
CC L-GLUTAMATE.
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO CLASS-I OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC AMINOTRANSFERASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL: D64001; BAA10261.1;
DR InterPro; IPR001176;
DR Pfam; PF00155; aminotran.1; 1.
DR PRINTS; PR00753; ACCSYNTHASE.
DR PROSITE; PS00105; AA_TRANSFER_CLASS_1; 1.
```


RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schütz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Sleeth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.,
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana";
RL Nature 402:769-777(1999).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=94368830; PubMed=8086440;
RA Horn M.A., Walker J.C.;
RT "Biochemical properties of the autophosphorylation of RLK5, a
receptor-like protein kinase from Arabidopsis thaliana";
RL Biochim. Biophys. Acta 1208:65-74(1994).
CC -!- COPACTOR: HAVE SIGNIFICANTLY GREATER ACTIVITY IN THE PRESENCE OF
MN2+ THAN MG2+.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN ROOTS AND ROSETTES.
CC -!- PTM: AUTOPHOSPHORYLATED ON SERINE AND THREONINE RESIDUES.
CC -!- SIMILARITY: TO PROTEIN KINASES IN THE C-TERMINUS. BUT DOES NOT
SEEM TO HAVE CONSERVED A KINASE ACTIVITY.
CC -!- SIMILARITY: CONTAINS 18 LEUCINE-RICH REPEATS (LRR).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84660; AAA32859.1; -
DR EMBL; AL021749; CAAL6889.1; -
DR EMBL; AL161572; CAB79651.1; -
DR HSP; P00523; 2PK.
DR InterPro; IPR000719; -
DR InterPro; IPR001611; -
DR InterPro; IPR002290; -
DR Pfam; PF00560; LRR; 14.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00019; LEUKICHRPT.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase: Serine/threonine-protein kinase; ATP-binding;
KW Transmembrane; Glycoprotein; Phosphorylation; Leucine-rich repeat;
KW Repeat; Signal.
FT SIGNAL 1 14
FT CHAIN 15 999
FT DOMAIN 15 621
FT TRANSMEM 622 641
FT DOMAIN 642 999
FT REPEAT 88 112
FT REPEAT 114 138
FT REPEAT 138 161
FT REPEAT 163 186
FT REPEAT 188 211
FT REPEAT 235 261
FT REPEAT 263 283
FT REPEAT 283 306
FT REPEAT 306 330
FT REPEAT 332 353
FT REPEAT 354 378
POTENTIAL.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
LRR 1.
LRR 2.
LRR 3.
LRR 4.
LRR 5.
LRR 6.
LRR 7.
LRR 8.
LRR 9.
LRR 10.
LRR 11.

FT REPEAT 402 426
FT REPEAT 426 450
FT REPEAT 452 474
FT REPEAT 498 522
FT REPEAT 524 547
FT REPEAT 549 567
FT REPEAT 567 592
FT DOMAIN 683 968
FT CARBOHYD 98 98
FT CARBOHYD 102 102
FT CARBOHYD 150 150
FT CARBOHYD 185 185
FT CARBOHYD 210 210
FT CARBOHYD 269 269
FT CARBOHYD 282 282
FT CARBOHYD 452 452
FT CARBOHYD 576 576
FT NP_BIND 689 697
FT BINDING 711 711
FT ACT_SITE 819 819
FT MUTAGEN 711 711
SQ SEQUENCE 999 AA; 109095 MW; F5793DB899EA0C6A7 CRC64;
K->E: LOSS OF CATALYTIC ACTIVITY.
BY SIMILARITY.
ATP (BY SIMILARITY).
BY SIMILARITY.
K->E: LOSS OF CATALYTIC ACTIVITY.
Query Match 55.9%; Score 38; DB 1; Length 999;
Best Local Similarity 70.0%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 GLICPDPRPL 11
DB 943 GLICTSPPL 952
RESULT 15
YGN3_YEAST
ID YGN3_YEAST STANDARD; PRT; 1364 AA.
AC P53125;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL 145.6 KDA PROTEIN IN RPL1B-CBG1 INTERGENIC REGION.
GN YGL133W OR G2842.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=S288C / FY1679;
RX MEDLINE=96437978; PubMed=8840506;
RA Escaribano V., Erasó P., Portillo F., Mazon M.J.;
RT "Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces
cerevisiae chromosome VII reveals SEC27, SSM1b, a putative
S-adenosylmethionine-dependent enzyme and six new open reading
frames";
RT Yeast 12:887-892(1996).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 272655; CA956844.1; -
DR SGD; S0003101; ITCL.
KW Hypothetical protein.
SQ SEQUENCE 1264 AA; 145642 MW; 45E4CF8835C7C746 CRC64;
Query Match 55.9%; Score 38; DB 1; Length 1264;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;

RA Laddison K.J., Specht C.A., Raper C.A., Kothe E.;
RT "the mating-type locus B alpha 1 of Schizosaccharomyces commune contains a
RL pheromone receptor gene and putative pheromone genes.";
EMBO J. 14:5271-5278(1995).
CC -!- FUNCTION: RECEPTOR FOR THE BAP2 PHEROMONE, A PRENYLATED MATING
CC FACTOR.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 4 OF G-PROTEIN COUPLED RECEPTORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X91168; CAA62595.1; -;
DR InterPro; IPR001499; -;
DR Pfam; PF02076; STE3; 1.
KW Transmembrane; G-protein coupled receptor; Pheromone response.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 36 56 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 110 130 POTENTIAL.
FT NON_TER 518 518
SQ SEQUENCE 518 AA; 57999 MW; AA635E46A1BF7C44 CRC64;

Query Match 54.4%; Score 37; DB 1; Length 518;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLLCPDPRP 10
I:| | | | |
DB 407 GVLVDPDPH 415

RESULT 19
GPV_MOUSE
ID GPV_MOUSE STANDARD; PRT; 567 AA.
AC O08742;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PLATELET GLYCOPROTEIN V PRECURSOR (GPV) (CD42D).
GN GP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Liver;
RX MEDLINE=97275136; PubMed=9129030;
RA Ravanat C., Morales M., Acosta D.O., Moog S., Schuhler S.,
RA Grunert P., Loew D., van Dorsselaer A., Cazenave J.-P., Lanza F.;
RT "Gene cloning of rat and mouse platelet glycoprotein V:
RT identification of megakaryocyte-specific promoters and demonstration
RT of functional thrombin cleavage.";
RL Blood 89:3253-3262(1997).
CC -!- FUNCTION: THE GPIIb-IIIa COMPLEX FUNCTIONS AS THE VON WILLEBRAND
CC FACTOR RECEPTOR AND MEDIATES VON WILLEBRAND FACTOR-DEPENDENT
CC PLATELET ADHESION TO BLOOD VESSELS. THE ADHESION OF PLATELETS TO
CC INJURED VASCULAR SURFACES IN THE ARTERIAL CIRCULATION IS A
CC CRITICAL INITIATING EVENT IN HEMOSTASIS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 14 LEUCINE-RICH REPEATS (LRR).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z69595; CAA93441.1; -;
DR MGD; MGI:1096363; GP5.
DR InterPro; IPR000483; -;
DR InterPro; IPR001611; -;
DR Pfam; PF00560; LRR; 13.
DR Pfam; PF01463; LRRCT; 1.
DR PRINTS; PRO0019; LEURICHRPT.
KW Platelet; Transmembrane; Glycoprotein; Blood coagulation;
KW Repeat; Leucine-rich repeat; Cell adhesion; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 567 PLATELET GLYCOPROTEIN V.
FT DOMAIN 17 522 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 523 543 POTENTIAL.
FT DOMAIN 544 567 CYTOPLASMIC (POTENTIAL).
FT REPEAT 73 96 LRR 1.
FT REPEAT 97 120 LRR 2.
FT REPEAT 122 144 LRR 3.
FT REPEAT 145 168 LRR 4.
FT REPEAT 170 192 LRR 5.
FT REPEAT 217 240 LRR 6.
FT REPEAT 241 264 LRR 7.
FT REPEAT 265 288 LRR 8.
FT REPEAT 289 312 LRR 9.
FT REPEAT 314 337 LRR 10.
FT REPEAT 338 361 LRR 11.
FT REPEAT 362 385 LRR 12.
FT REPEAT 386 409 LRR 13.
FT REPEAT 409 518 LRR 14.
FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 567 AA; 63467 MW; C48643AA73967A7D CRC64;

Query Match 54.4%; Score 37; DB 1; Length 567;
Best Local Similarity 85.7%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 CPDPRPL 11
I:| | | | |
DB 472 CPDPRSL 478

RESULT 20
LEPA_MYCTU
ID LEPA_MYCTU STANDARD; PRT; 653 AA.
AC P71739;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GTP-BINDING PROTEIN LEPA.
GN LEPA OR RV2404C OR WTCY253.16.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Ousell M.A., Rajandream M.A., Rogers J.
RA Rutter S., Saeger K., Skellton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrall B.G.:
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: Z81368; CAB03723.1;
CC HSSP: P13551; LELO.
CC Tuberculin: RV2404C; -.
CC Pfam: PF00009; GTP_EFTU; 1.
CC PROSITE: PS00301; EFACOR_GTP; 1.
CC GTP-binding. 59 66 GTP (BY SIMILARITY).
CC NP_BIND 124 128 GTP (BY SIMILARITY).
CC NP_BIND 178 181 GTP (BY SIMILARITY).
CC NP_BIND 633 AA: 72395 MW; DA4AFEL0E6C25755 CRC64;
CC SEQUENCE
Query Match 54.4%; Score 37; DB 1; Length 653;
Best Local Similarity 45.5%; Pred. No. 86;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 2 GILCPDPRPLE 12
DB 285 GIVSPKPCPE 295

RESULT 21
ID NCYC_SYPN7 STANDARD; PRT; 659 AA.
AC P38045;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE NITRATE TRANSPORT ATP-BINDING PROTEIN NRTC.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93173091; PubMed=8437564;
RA "Mata L.F., Andriess X., Hirano A.:
RT "Nitrate transport in the cyanobacterium Synechococcus sp. PCC7942."
RL Mol. Gen. Genet. 236:193-202(1993).
CC -1- FUNCTION: PROBABLY PART OF A HIGH-AFFINITY BINDING-PROTEIN-
CC DEPENDENT TRANSPORT SYSTEM FOR NITRATE. PROBABLY RESPONSIBLE FOR
CC ENERGY COUPLING TO THE TRANSPORT SYSTEM.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (POTENTIAL)
CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS).
CC -1- SIMILARITY: SOME, IN THE C-TERMINAL DOMAIN TO NRTA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch.
CC -----

CC EMBL: X61625; CAA43811.1;
CC PIR: S30893; S30893.
CC InterPro: IPR001617;
CC Pfam: PF00005; ABC_tran; 1.
CC PROSITE: PS00311; ABC_TRANSPORTER; 1.
CC TRANSPORT; ATP-binding; Membrane; Nitrate assimilation.
CC ACT_SITE 278
CC DOMAIN 255 278 LINKER.
CC NP_BIND 42 49 NTRA-LIKE.
CC NP_BIND 279 659
CC SEQUENCE 659 AA: 72346 MW; 1D445792A5A76134 CRC64;
Query Match 54.4%; Score 37; DB 1; Length 659;
Best Local Similarity 77.8%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGLLCPDPR 9
DB 506 AGELCDPDR 514

RESULT 22
ID UNG_CHLTR STANDARD; PRT; 229 AA.
AC O84613; 2000 (Rel. 30, Created)
DT 30-MAY-2000 (Rel. 30, Last annotation update)
DT 30-MAY-2000 (Rel. 30, Last annotation update)
DE URACIL-DNA GLYCOSYLASE (EC 3.2.2.-) (UDG).
CC UNG OR CT607.
CC Chlamydia trachomatis.
CC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
CC NP_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=D/UV-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: EXCISES URACIL RESIDUES FROM THE DNA WHICH CAN ARISE
CC AS A RESULT OF MISINCORPORATION OF DUMP RESIDUES BY DNA
CC POLYMERASE OR DUE TO DEAMINATION OF CYTOSINE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE URACIL-DNA GLYCOSYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: AB001331; AAC68210.1;
CC HSSP: P12295; LEUT.
CC InterPro: IPR002043;
CC Pfam: PF00130; UDNA_GLYCOSYLASE; 1.
CC PROSITE: PS00130; UDNA_GLYCOSYLASE; 1.
CC DNA repair; Hydrolase; Glycosidase.
CC ACT_SITE 70 70
CC SEQUENCE 229 AA: 26002 MW; 9AD49846F8DC3B CRC64;
Query Match 52.9%; Score 36; DB 1; Length 229;
Best Local Similarity 63.6%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 AGLCPDPRPL 11
DB 187 AVLACPHPSPL 197

RESULT 23
IBP3_BOVIN
ID IBP3_BOVIN STANDARD; PRT; 291 AA.
AC P20959;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 3 PRECURSOR (IGFBP-3)
DE (IBP-3) (IGF-BINDING PROTEIN 3).
GN IGFBP3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91282738; PubMed=1711841;
RA Spratt S.K., Tatsuno G.P., Sommer A.;
RT "Cloning and characterization of bovine insulin-like growth factor
binding protein-3 (IGFBP-3).";
RL Biochem. Biophys. Res. Commun. 177:1025-1032(1991).
RN [2]
RP SEQUENCE OF 28-52.
RX MEDLINE=91065246; PubMed=1701128;
RA Conover C.A., Ronk M., Lombana F., Powell D.R.;
RT "Structural and biological characterization of bovine insulin-like
growth factor binding protein-3.";
RL Endocrinology 127:2795-2803(1990).
RN [3]
RP SEQUENCE OF 194-266 FROM N.A.
RX MEDLINE=98030060; PubMed=9363609;
RA Maciulla J.H., Zhang H.M., Denise S.K.;
RT "A novel polymorphism in the bovine insulin-like growth factor
binding protein-3 (IGFBP3) gene.";
RL Anim. Genet. 28:375-375(1997).
CC -|- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs
AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE
INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.
CC -|- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH
IGF-I OR IGF-II AND A 85 KDA GLYCOPROTEIN (ALS).
CC -|- SUBCELLULAR LOCATION: SECRETED.
CC -|- TISSUE SPECIFICITY: PLASMA; EXPRESSED BY MOST TISSUES.
CC -|- MISCELLANEOUS: BINDS IGF-II MORE THAN IGF-I.
CC -|- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DB EMBL; M76478; AAA30582.1; -
DR EMBL; U83465; AAB41430.1; -
DR PIR; JN0064; JN0064.
DR InterPro; IPR000716; -
DR InterPro; IPR000867; -
DR Pfam; PF00219; IGFBP; 1.
DR Pfam; PF00086; thyroglobulin_1; 1.
DR PROSITE; PS00222; IGF BINDING; 1.
DR PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW Growth factor binding; Signal; Glycoprotein.
FT SIGNAL 1 27

QY 1 AGLICP-----DPRPLE 12
DB 92 SGLRCQPPGDPRLQ 107

Query Match 52.9%; Score 36; DB 1; Length 291;
Best Local Similarity 50.0%; Pred. No. 59;
Matches 8; Conservative 2; Mismatches 2; Indels 4; Gaps 1;

QY 1 AGLICP-----DPRPLE 12
DB 92 SGLRCQPPGDPRLQ 107

RESULT 24
TCBF_PSESQ STANDARD; PRT; 352 AA.
ID TCBF_PSESQ STANDARD; PRT; 352 AA.
AC P27101;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MALEYLACETATE REDUCTASE (EC 1.3.1.32).
GN TCBF.
OS Pseudomonas sp. (strain P51).
OG Plasmid pp51.
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae.
OX NCBI_TaxID=65067;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91193197; PubMed=2013566;
RA van der Meer J.R., Eggen R.I., Zehnder A.J., de Vos W.M.;
RT "Sequence analysis of the Pseudomonas sp. strain P51 tcb gene
cluster, which encodes metabolism of chlorinated catechols: evidence
for specialization of catechol 1,2-dioxygenases for chlorinated
substrates.";
RL J. Bacteriol. 173:2425-2434(1991).
CC -|- CATALYTIC ACTIVITY: 3-OXOADIPATE + NAD(P)(+) = 2-MALEYLACETATE +
NAD(P)H.
CC -|- PATHWAY: 3-CHLOROCATECHOL DEGRADATION (BETA-KETOADIPATE PATHWAY).
CC THIS PATHWAY SERVES A VITAL ROLE IN THE BIODEGRADATION OF TOXIC
AROMATIC COMPOUNDS INTRODUCED IN THE ENVIRONMENT BOTH AS NATURAL
PRODUCTS AND AS INDUSTRIAL EFFLUENT.
CC -|- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
FAMILY.
CC -|- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A TRANS-
DIENELACTONISOMERASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; M57629; AAD13629.1; -
DR PIR; E43673; E43673.
DR InterPro; IPR001670; -
DR Pfam; PF00465; Fe-ADH; 2.
DR PROSITE; PS00913; ADH_IRON_1; FALSE_NEG.
DR PROSITE; PS00060; ADH_IRON_2; FALSE_NEG.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD; Plasmid.
SQ SEQUENCE 352 AA; 37499 MW; C72A9D2671FFFB3 CRC64;

```

```

Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 0; Indels 0; Gaps 0;

QY 6 PDRPLE 12
DB 327 PNRPLE 333

RESULT 25
TFTE_BURCE
ID TFTE_BURCE STANDARD; PRT; 352 AA.
AC Q45072;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 01-AUG-1992 (Rel. 23, Last annotation update)
GN TFTEALACETATE REDUCTASE (EC 1.3.1.32).
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Burkholderia; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=292;
RP SEQUENCE FROM N.A.
RA MEDLINE=95266809; PubMed=7538273;
RX Daubaras D.L., Hersberger C.D., Kitano K., Chakrabarty A.M.;
RT "Sequence analysis of a gene cluster involved in metabolism of 2,4,5-
trichlorophenoxyacetic acid by Burkholderia cepacia AC1100.";
RN J. Bacteriol. 172:2351-2359(1990).
KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD; Plasmid.
SQ SEQUENCE 352 AA; 36823 MW; FD7F0F8959357044 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 352;
Best Local Similarity 85.7%; Pred. No. 70;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 PDRPLE 12
DB 328 PNRPLE 334

RESULT 26
TFTE1_ALCEU
ID TFTE1_ALCEU STANDARD; PRT; 354 AA.
AC P27137;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 01-AUG-1992 (Rel. 23, Last annotation update)
GN 2,4,5-TRICHLOROPHENOXYACETATE REDUCTASE I (EC 1.3.1.32).
OS Alcaligenes eutrophus (Ralstonia eutropha).
OC Alcaligenes eutrophus (Ralstonia eutropha).
OX Plasmid pJP4.

```

```

Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
OX NCBI_TaxID=510;
RP SEQUENCE FROM N.A.
RA STRAIN=JMI13489; PubMed=2185214;
RX Becking E.J., Gordon M.P., Caceres O., Lurquin P.F.;
RT "Organization and sequence analysis of the 2,4-dichlorophenol
hydroxylase and dichlorocatechol oxidative operons of plasmid pJP4.";
RN J. Bacteriol. 172:2351-2359(1990).
KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD; Plasmid.
SQ SEQUENCE 354 AA; 37899 MW; 7D8C001B5B13376 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 354;
Best Local Similarity 85.7%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 PDRPLE 12
DB 328 PNRPLE 334

RESULT 27
HXB2_HUMAN
ID HXB2_HUMAN STANDARD; PRT; 356 AA.
AC P14652; P17485; P10913;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DE 01-APR-1990 (Rel. 14, Last annotation update)
GN HOMEOBOX PROTEIN HOXB-2 (HOX-2H) (HOX-2.8) (K8).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RA MEDLINE=90098876; PubMed=2574852;
RX Acampora D., D'Esposito M., Faella A., Pannese M., Migliacchio E.,
RA Morelli F., Stornaiuolo A., Nigro V., Simeone A., Boncinelli E.;
RT "The human HOX gene family.";
RN Nucleic Acids Res. 17:10385-10402(1989).
KW Homeobox proteins; Hox genes; HOX-2H; HOX-2.8; HOX-2.9; HOX-2.10;
SQ SEQUENCE OF 132-208 FROM N.A.
OX Tissue=Placenta;
RX MEDLINE=89378558; PubMed=2570724;
RX Giampaolo A., Acampora D., Zappavigna V., Pannese M.,

```

RA D'Esposito M., Care A., Faiella A., Stornaiuolo A., Russo G.,
RA Simeone A., Boncinelli E., Peschle C.;
RT "Differential expression of human HOX-2 genes along the anterior-
RT posterior axis in embryonic central nervous system.";
RL Differentiation 40:191-197(1989).
RN [3]
RP SEQUENCE OF 143-208 FROM N.A.
RX MEDLINE=90215256; PubMed=2576652;
RA Boncinelli E., Acampora D., Pannese M., D'Esposito M., Somma R.,
RA Gaetano G., Stornaiuolo A., Cafiero M., Faiella A., Simeone A.;
RT "Organization of human class I homeobox genes.";
RL Genome 31:745-756(1989).
RN [4]
RP SEQUENCE OF 143-202 FROM N.A.
RX MEDLINE=88329001; PubMed=2501346;
RA Kongsuwan K., Webb E., Housiaux P., Adams J.M.;
RT "Expression of multiple homeobox genes within diverse mammalian
RT haemopoietic lineages.";
RL EMBO J. 7:2131-2138(1988).
RN [5]
RP SEQUENCE OF 1-42 FROM N.A.
RX MEDLINE=95181447; PubMed=7876223;
RA Vieille-Grosjean I., Huber P.;
RT "Transcription factor GATA-1 regulates human HOXB2 gene expression in
RT erythroid cells.";
RL J. Biol. Chem. 270:4544-4550(1995).
CC -!- FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
CC A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
CC SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE EMBRYOS AND FETUSES AT
CC 5-9 WEEKS FROM CONCEPTION.
CC -!- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.
CC PROBOSCIPEDIA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X16665; CAA34655.1; -;
DR EMBL; X16176; CAA34298.1; -;
DR EMBL; X14571; CAA32709.1; -;
DR EMBL; X78978; CAA55581.1; -;
DR PIR; S07542; WJHU2H.
DR PIR; E37042; E37042.
DR HSP; P02833; ISAN.
DR MIM; 142967; -;
DR InterPro; IPR001356; -;
DR InterPro; IPR001827; -;
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR PRINTS; PR00025; ANTENNAPEDIA.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00032; ANTENNAPEDIA; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 94 99 ANTP-TYPE HEXAPEPTIDE.
FT DNA_BIND 143 202 HOMEBOX.
FT CONFLICT 136 138 PEA -> RRL (IN REF. 2).
SQ SEQUENCE 356 AA; 37913 MW; D7ACA0E3D5BEBF9 CRC64;

Query Match 52.9%; Score 36; DB 1; Length 356;
Best Local Similarity 70.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GLLCPDRPL 11
| | | | |

DB 254 GALSADPRPL 263
RESULT 28
CA26_CHICK STANDARD; PRT; 1022 AA.
ID CA26_CHICK
AC P15988;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE COLLAGEN ALPHA 2(VI) CHAIN PRECURSOR.
OS COL6A2.
GN Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89305506; PubMed=2787244;
RA Koller E., Winterhalter K.H., Trueb B.;
RT "The globular domains of type VI collagen are related to the
RT collagen-binding domains of cartilage matrix protein and von
RT Willebrand factor.";
RL EMBO J. 8:1073-1077(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91200044; PubMed=2015818;
RA Hayman A.R., Koppel J., Trueb B.;
RT "Complete structure of the chicken alpha 2(VI) collagen gene.";
RL Eur. J. Biochem. 197:177-184(1991).
RN [3]
RP SEQUENCE OF 8-38 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=91187664; PubMed=2011522;
RA Koller E., Hayman A.R., Trueb B.;
RT "The promoter of the chicken alpha 2(VI) collagen gene has features
RT characteristic of house-keeping genes and of proto-oncogenes.";
RL Nucleic Acids Res. 19:485-491(1991).
CC -!- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(VI),
CC ALPHA 2(VI), AND ALPHA 3(VI).
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15041; CAA33144.1; -;
DR EMBL; X56659; CAA39982.1; -;
DR EMBL; X56595; CAA39933.1; -;
DR PIR; S04111; S04111.
DR PIR; S23378; S23378.
DR InterPro; IPR000087; -;
DR InterPro; IPR002035; -;
DR Pfam; PF01391; Collagen; 5.
DR Pfam; PF00092; vwa; 3.
DR PRINTS; PR00453; VWFA_DOMAIN.
DR PROSITE; PS0234; VWFA_DOMAIN; 3.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Cell adhesion; Collagen; Signal.
FT SIGNAL 1 27
FT CHAIN 28 1022 COLLAGEN ALPHA 2(VI) CHAIN.
FT DOMAIN 28 255 NONHELICAL REGION.
FT DOMAIN 256 590 TRIPLE-HELICAL REGION.
FT DOMAIN 591 1022 NONHELICAL REGION.
FT SITE 514 519 INTERRUPTION IN COLLAGENOUS REGION.
FT DOMAIN 44 168 VWFA 1.

DR DOMAIN 613 738 WIFA 2.
 DR SITE 933 957 WIFA 3.
 FT SITE 348 350 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 366 368 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 426 428 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 444 446 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 465 467 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 483 485 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 498 500 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 897 897 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1022 AA; 109176 MW; 3194CFD1475AE893 CRC64.

 Query Match 52.9%; Score 36; DB 1; Length 1022;
 Best Local Similarity 71.4%; Pred. No. 1.9e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 3 LCPDPR 9
 DB 804 MLCDPDPO 810

 RESULT 29
 YDY_MOUSE MOUSE STANDARD; PRT: 1212 AA.
 AC DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE UBQUITOUSLY TRANSCRIBED Y CHROMOSOME TETRATRIPEPTIDE REPEAT PROTEIN
 DE UBQUITOUSLY TRANSCRIBED TPR PROTEIN ON THE Y CHROMOSOME (MALE-
 CN UTILITY HISTOCOMPATIBILITY ANTIGEN H-YDB).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K111; TISSUE=Testis;
 RX MEDLINE=98409500; PubMed=9736773;
 RA Mazyrat S., Saut N., Sargent C.A., Grimmond S., Longepied G.,
 Ehrmann I.E., Ellis P.S., Greenfield A., Affara N.A., Mitchell M.J.;
 FT "The mouse Y chromosome interval necessary for spermatogonial
 proliferation is gene dense with syntenic homology to the human AZFa
 Hum. Mol. Genet. 7:1713-1724(1998).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS;
 RX MEDLINE=9709462; PubMed=8944031;
 RA Sienfeldt A., Scott P., Pennisi D., Ehrmann I., Ellis P., Cooper L.,
 RT "An H-Ydb epitope is encoded by a novel mouse Y chromosome gene."
 RL Nat. Genet. 14:474-478(1996).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -!- SIMILARITY: CONTAINS 6 TPR REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL database.
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 DR EMBL: AF057367; AAC67385.1;
 DR MGD: MGI:894810; Uty.

DR INTERPTG: IPR001440; -;
 DR Pfam: PF00515; TPR; Nuclear protein.
 KW Repeat; TPR repeat; 119
 FT DOMAIN 91 119 TPR 1.
 FT DOMAIN 128 156 TPR 2.
 FT DOMAIN 165 193 TPR 3.
 FT DOMAIN 203 231 TPR 4.
 FT DOMAIN 231 269 TPR 5.
 FT DOMAIN 269 307 TPR 6.
 FT CONFLICT 1069 1069 E -> Q (IN REF. 2).
 FT CONFLICT 1149 1212 YKMDLLTVYDFTLAPSLSSAS -> STRDLLPQHLRQC
 FT HLOQPTDKAAALIEHLTSGSDMH (IN REF. 2).
 SQ SEQUENCE 1212 AA; 136736 MW; 2AEIA816FD6ACB5 CRC64;

 Query Match 52.9%; Score 36; DB 1; Length 1212;
 Best Local Similarity 71.4%; Pred. No. 2.2e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 4 LCPDPR 10
 DB 729 ICPDPRP 735

 RESULT 30
 YDM5_SCHPO STANDARD; PRT: 1337 AA.
 ID DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHETICAL 150.8 KDA PROTEIN C57A7.05 IN CHROMOSOME 1.
 GN SPAC57A7.05.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Schizosaccharomycetes; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972; Churche C.M., Wood V., Barrell B.G., Rajandream M.A.;
 RX MEDLINE=9709462; PubMed=8944031;
 RA Sienfeldt A., Scott P., Pennisi D., Ehrmann I., Ellis P., Cooper L.,
 RT "An H-Ydb epitope is encoded by a novel mouse Y chromosome gene."
 RL Nat. Genet. 14:474-478(1996).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -!- SIMILARITY: SOME, TO YEAST YDL231C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL database.
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 DR EMBL: Z95396; CAB08763.1;
 DR MGD: MGI:894810; Uty.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 209 229 POTENTIAL.
 FT TRANSMEM 241 261 POTENTIAL.
 FT TRANSMEM 267 287 POTENTIAL.
 FT TRANSMEM 328 348 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 481 500 POTENTIAL.
 FT TRANSMEM 917 937 POTENTIAL.
 FT TRANSMEM 975 995 POTENTIAL.
 FT TRANSMEM 997 1017 POTENTIAL.
 FT TRANSMEM 1021 1041 POTENTIAL.
 FT TRANSMEM 1066 1086 POTENTIAL.
 FT TRANSMEM 1275 1293 POTENTIAL.
 SQ SEQUENCE 1337 AA; 150848 MW; A549BC8E0D08791 CRC64;

 Query Match 52.9%; Score 36; DB 1; Length 1337;

Query Match 52.2%; Score 35.5; DB 1; Length 501;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY 4 LCPD---PRPL 11

DB 394 LCPDCRAPRI 404

RESULT 35

PLF4_SHEEP

ID PLF4_SHEEP

AC P30035;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DE 01-OCT-2000 (Rel. 40, Last annotation update)

DE PLATELET FACTOR 4 (PF-4).

GN SCYB4 OR PF4.

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Caprinae; Ovis.

OX NCBI_TaxID=9940;

[1]

SEQUENCE.

RA MEDLINE=92160127; PubMed=1788836;

RA Shigeta O., Lu W., Holt J.C., Edmunds L.H. Jr., Niewiarowski S.;

RT "Ovine platelet factor 4: purification, amino acid sequence,

RT radioimmunoassay and comparison with platelet factor 4 of other

RT species.";

RL Thromb. Res. 64:509-520(1991).

CC -!- FUNCTION: PLATELET FACTOR 4, NONCOVALENTLY BOUND TO A PROTEOGLYCAN

CC MOLECULE, IS RELEASED DURING PLATELET AGGREGATION. PF4 NEUTRALIZES

CC THE ANTICOAGULANT EFFECT OF HEPARIN BECAUSE IT BINDS MORE STRONGLY

CC TO HEPARIN THAN TO THE CHONDROITIN-4-SULFATE CHAINS OF THE CARRIER

CC MOLECULE. CHEMOTACTIC FOR NEUTROPHILS AND MONOCYTES.

CC -!- SUBUNIT: HOMOTETRAMER.

CC -!- SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE

CC C-X-C) (CHEMOKINE CXCL).

DR HSP: P02777; 1PLF.

DR InterPro: IPR001089; -.

DR Pfam: PF00048; IL8; 1.

DR PRINTS: PR00437; SMALLCYTKCMX.

DR PROSITE: PS00471; SMALL_CYTOKINES_CXC; 1.

KW Cytokine; Platelet; Proteoglycan; Heparin-binding; Chemotaxis.

FT DISULFID 25 51 BY SIMILARITY.

FT DISULFID 27 67 BY SIMILARITY.

SQ SEQUENCE 85 AA; 9129 MW; 925A76512095ECF7 CRC64;

Query Match

Best Local Similarity 51.5%; Score 35; DB 1; Length 85;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGLLCPDPR 9

DB 47 AGLHCPSPQ 55

RESULT 36

SR19_ORYSA

ID SR19_ORYSA

AC P49964;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DE 01-OCT-1996 (Rel. 34, Last annotation update)

DE SIGNAL RECOGNITION PARTICLE 19 KDA PROTEIN (SRP19).

GN SRP19.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae;

OC Oryza.

OX NCBI_TaxID=4530;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE / JAPONICA; TISSUE=Callus;

RA Zwiab C.W., Black S.D.;

RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: SIGNAL-RECOGNITION-PARTICLE ASSEMBLY, BINDS DIRECTLY

CC TO 7S RNA AND MEDIATES BINDING OF THE 54 KDA SUBUNIT OF THE SRP.

CC (BY SIMILARITY).

CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 7S RNA MOLECULE

CC OF 300 NUCLEOTIDES AND SIX PROTEIN SUBUNITS: SRP72, SRP68, SRP54,

CC SRP19, SRP14 AND SRP9 (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -!- SIMILARITY: BELONGS TO THE SRP19 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

CC -----

CC EMBL; U19030; AAB65810.1; -.

DR InterPro: IPR002778; -.

DR Pfam: PF01922; SRP19; 1.

KW Signal recognition particle; RNA-binding; Ribonucleoprotein.

FT DOMAIN 127 136 BASIC REGION, POTENTIALLY INVOLVED IN

FT RNA-BINDING.

SQ SEQUENCE 136 AA; 14922 MW; 65D3B91967F3D638 CRC64;

Query Match

Best Local Similarity 51.5%; Score 35; DB 1; Length 136;

Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGLLCPDPRPLE 12

DB 35 SGRACPDPTCVE 46

RESULT 37

ID IBP3_PIG

AC PI6611;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 3 (IGFBP-3) (IBP-3) (IGF-

DE BINDING PROTEIN 3).

GN IGFBP3.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

[1]

SEQUENCE FROM N.A.

RP MEDLINE=90130475; PubMed=1688850;

RA Shimazaki S., Shimonaka M., Ui M., Inouye S., Shibata F., Ling N.;

RT "Structural characterization of a follicle-stimulating hormone action

RT inhibitor in porcine ovarian follicular fluid. Its identification as

RT the insulin-like growth factor-binding protein.";

RL J. Biol. Chem. 265:2198-2202(1990).

[2]

SEQUENCE OF 1-15.

RP MEDLINE=92109718; PubMed=1722398;

RA Coleman M.E., Pan Y.-C.E., Etherton T.D.;

RT "Identification and NH2-terminal amino acid sequence of three

RT insulin-like growth factor-binding proteins in porcine serum.";

RL Biochem. Biophys. Res. Commun. 181:1131-1136(1991).

CC -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS

CC AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH

CC PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE

CC INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.

CC -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KDA WITH

CC IGF-I OR IGF-II AND A 85 KDA GLYCOPROTEIN (ALS).
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- MISCELLANEOUS: BINDS IGF-II MORE THAN IGF-I.
 CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
 CC PROTEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J05228; AAA31054.1; -
 CC PIR: A35037; A35037.
 CC PIR: JH0516; JH0516.
 CC HSSP: P17494; IKST.
 CC InterPro: IPR000716; -
 CC Pfam: PF00001; IGF1P.1.
 CC Pfam: PF00219; IGF1P.1.
 CC Pfam: PF00086; thyroglobulin_1; 1.
 CC PROSITE: PS00222; IGF-BINDING; 1.
 CC Growth factor binding; thyroglobulin_1; 1.
 CC CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC DOMAIN 211 260 THYROGLOBULIN TYPE I.
 CC CONFLICT 5 6 VG -> A (IN REF. 2).
 CC SEQUENCE 266 AA; 28910 MW; CB65809ABA71D670 CRC64;
 DR PIR: A35037; A35037.
 DR PIR: JH0516; JH0516.
 DR HSSP: P17494; IKST.
 DR InterPro: IPR000716; -
 DR Pfam: PF00001; IGF1P.1.
 DR Pfam: PF00219; IGF1P.1.
 DR Pfam: PF00086; thyroglobulin_1; 1.
 DR PROSITE: PS00222; IGF-BINDING; 1.
 DR Growth factor binding; thyroglobulin_1; 1.
 DR CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DR DOMAIN 211 260 THYROGLOBULIN TYPE I.
 DR CONFLICT 5 6 VG -> A (IN REF. 2).
 DR SEQUENCE 266 AA; 28910 MW; CB65809ABA71D670 CRC64;
 KW Growth factor binding; thyroglobulin_1; 1.
 KW CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW DOMAIN 211 260 THYROGLOBULIN TYPE I.
 KW CONFLICT 5 6 VG -> A (IN REF. 2).
 KW SEQUENCE 266 AA; 28910 MW; CB65809ABA71D670 CRC64;
 QY 1 AGILCP----DPRLE 12
 DB 65 AGLRCPPPGPEPRPLQ 80
 RESULT 38
 ID CYSH_PSEAE STANDARD; PRT; 267 AA.
 AC 005927; 1997 (Rel. 35, Created)
 DT 01-AUG-1997 (Rel. 35, Last sequence update)
 DE PHOSPHADENOSINE PHOSPHOSULFATE REDUCTASE (EC 1.8.99.4) (PAPS
 DE REDUCTASE, THIOREDOKIN DEPENDENT) (PAPS REDUCTASE) (3'-
 DE PHOSPHADENYL SULFATE REDUCTASE) (PAPS SULFOTRANSFERASE).
 GN CYSH OR P1756.
 GN Pseudomonas aeruginosa.
 OC Bacteroidetes; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OC NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CHA161814; PubMed=9218775;
 RA Delic-Attree I, Toussaint B., Garin J., Vignais P.M.;
 RT "Cloning, sequence and mutagenesis of the structural gene of
 RT Pseudomonas aeruginosa CysB, which can activate algd transcription.";
 RL Mol. Microbiol. 24:1275-1284(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000)
 CC -1- CATALYTIC ACTIVITY: 5'-PHOSPHADENOSINE 3'-PHOSPHOSULFATE + REDUCED
 CC THIOREDOKIN = PHOSPHADENOSINE PHOSPHATE + OXIDIZED THIOREDOKIN +
 CC SULFITE.
 CC -1- PATHWAY: THIRD STEP IN THE SULFATE ACTIVATION PATHWAY IN THE
 CC REDUCTIVE BRANCH OF THE CYSTEINE BIOSYNTHETIC PATHWAY.
 CC -1- SUBCELLULAR LOCATION: CYTOSOL.
 CC -1- SIMILARITY: BELONGS TO THE PAPS REDUCTASE FAMILY. CYSH SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U95379; AAB53743.1; -
 CC EMBL: AEO04601; AAG05145.1; -
 CC InterPro: IPR002500; reduct; 1.
 CC Pfam: PF00001; IGF1P.1.
 CC Pfam: PF00086; thyroglobulin_1; 1.
 CC PROSITE: PS00222; IGF-BINDING; 1.
 CC Growth factor binding; thyroglobulin_1; 1.
 CC CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC DOMAIN 211 260 THYROGLOBULIN TYPE I.
 CC CONFLICT 5 6 VG -> A (IN REF. 2).
 CC SEQUENCE 267 AA; 30215 MW; E9557E1970F21049 CRC64;
 KW Growth factor binding; thyroglobulin_1; 1.
 KW CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 KW DOMAIN 211 260 THYROGLOBULIN TYPE I.
 KW CONFLICT 5 6 VG -> A (IN REF. 2).
 KW SEQUENCE 267 AA; 30215 MW; E9557E1970F21049 CRC64;
 QY 3 LKCPDPRPLE 12
 DB 111 VLSPDPRLE 120
 RESULT 39
 ID YPOQ_KLEPN STANDARD; PRT; 271 AA.
 AC P27509;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DE HYPOTHETICAL PROTEIN IN POQA 5'REGION (ORF X) (FRAGMENT).
 OS Klebsiella pneumoniae.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Klebsiella.
 OC NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=92212293; PubMed=1313537;
 RA Meulenbergh J.J.M., Sellink E., Riegman N.H., Postma P.W.;
 RT "Nucleotide sequence and structure of the Klebsiella pneumoniae pqq
 RT operon.";
 RL Mol. Microbiol. 23:284-294(1992).
 RN [2]
 RP FUNCTION: NOT KNOWN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M19.
 CC -1- SIMILARITY: TO AN ORF IN THE 3'REGION OF POQ-III IN
 CC A.CALCOACETICUS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

```
CC -----
DR EMBL; X58778; CAA41578.1; -
DR PIR; S20452; S20452.
DR MEROPS; M19.003; -
DR InterPro; IPR000180; -
DR PROSITE; PS00869; RENAL_DIPEPTIDASE; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 271 AA; 29484 MW; CB8AB11F3C8EC42E CRC64;

Query Match 51.5%; Score 35; DB 1; Length 271;
Best Local Similarity 53.6%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 AGLLCPDPRPL 11
DB 150 AHALCPQPRNL 160

RESULT 40
FCN3_HUMAN
ID FCN3_HUMAN STANDARD; PRT; 299 AA.
AC 075636;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FCN3
DE (COLLAGEN/FIBRINOGEN DOMAIN-CONTAINING LECTIN 3 P35) (HAKATA ANTIGEN).
GN FCN3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=98361935; PubMed=9694814;
RA Sugimoto R., Yae Y., Akaiwa M., Kitajima S., Shibata Y., Sato H.,
RA Hirata J., Okochi K., Izuhara K., Hamasaki N.;
RT "Cloning and characterization of the Hakata antigen, a member of the
RT ficollin/opsolin p35 lectin family.";
RL J. Biol. Chem. 273:20721-20727(1998).
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=99264375; PubMed=10330454;
RA Akaiwa M., Yae Y., Sugimoto R., Suzuki S.O., Iwaki T., Izuhara K.,
RA Hamasaki N.;
RT "Hakata antigen, a new member of the ficollin/opsolin p35 family, is a
RT novel human lectin secreted into bronchus/alveolus and bile.";
RL J. Histochem. Cytochem. 47:777-786(1999).
CC -!- FUNCTION: INVOLVED IN THE SERUM EXERTING LECTIN ACTIVITY. HAS
CC AFFINITY WITH GALNAC, GLCNAC, D-FUCOSE, AS MONO/OLIGOSACCHARIDE
CC AND LIPOPOLYSACCHARIDES FROM S.TYPHIMURIUM AND S.MINNESOTA.
CC -!- SUBUNIT: DISULFIDE-LINKED HOMOPOLYMER. MAY BE A OCTADECAMER
CC CONSISTING OF AN ELEMENTARY TRIMER UNIT. DOES NOT INTERACT WITH
CC FIBRONECTIN, ELASTIN OR ZYMOSAN.
CC -!- SUBCELLULAR LOCATION: SECRETED IN PLASMA, BRONCHUS, ALVEOLUS AND
CC BILE DUCT.
CC -!- TISSUE SPECIFICITY: LIVER AND LUNG. IN LIVER IT IS PRODUCED BY
CC BILE DUCT EPITHELIAL CELLS AND HEPATOCYTES. IN LUNG IT IS PRODUCED
CC BY BOTH CILIATED BRONCHIAL EPITHELIAL CELLS AND TYPE II ALVEOLAR
CC EPITHELIAL CELLS.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- DISEASE: ANTIGEN FOUND 14 TIMES MORE FREQUENTLY IN THE SERUM OF
CC PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS (SLE) THAN IN PATIENTS
CC WITH OTHER AUTOIMMUNE DISEASES.
CC -!- SIMILARITY: BELONGS TO THE FICOLLIN LECTIN FAMILY.
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 FIBRINOGEN C-TERMINAL DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D88587; BAA32277.1; -
DR MIM; 604973; -
DR HSP; P02671; 1FZD
DR InterPro; IPR000087; -
DR InterPro; IPR002181; -
DR Pfam; PF00147; fibrinogen_C; 1.
DR PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; 1.
KW Lectin; Collagen; Repeat; Glycoprotein; Signal; Multigene family;
KW Antigen; Hydroxylation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 299 FICOLLIN 3.
FT DOMAIN 48 80 COLLAGEN-LIKE.
FT DOMAIN 119 265 FIBRINOGEN C-TERMINAL.
FT MOD_RES 50 50 HYDROXYLATION.
FT MOD_RES 53 53 HYDROXYLATION.
FT MOD_RES 59 59 HYDROXYLATION.
FT MOD_RES 65 65 HYDROXYLATION.
FT MOD_RES 68 68 HYDROXYLATION.
FT MOD_RES 77 77 HYDROXYLATION.
FT CARBOHYD 189 189 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 299 AA; 32889 MW; 5CB8A7D3679FB264 CRC64;

Query Match 51.5%; Score 35; DB 1; Length 299;
Best Local Similarity 75.0%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 5 CPDPRPLE 12
DB 29 CPGPRELE 36

RESULT 41
RM04_YEAST
ID RM04_YEAST STANDARD; PRT; 319 AA.
AC P36517;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 60S RIBOSOMAL PROTEIN L4, MITOCHONDRIAL PRECURSOR (YML4).
GN MRPL4 OR YLR439W OR L9753.1.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=07173;
RA Graack H.-R.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Pavello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Jordis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevaskis E., Vignati D.,
RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 15-41.
RX MEDLINE=89078618; PubMed=3060376;
RA Graack H.-R., Grohmann L., Choli T.;
RT "Mitochondrial ribosomes of yeast: isolation of individual proteins
RT and N-terminal sequencing.";
```

RL FEBS Lett. 242:4-8(1988).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z30582; CAAB3057.1; -;
 DR EMBL; U21094; AAB67513.1; -;
 DR PIR; S26755; S26755.
 DR SCD; S0004431; MRP14
 RN Mitochondrial protein; 14 MITOCHONDRION. Transist peptide.
 FT TRANSIT
 FT CHAIN 15 319 60S RIBOSOMAL PROTEIN L4.
 FT CONFLICT 168 168 E -> K (IN REF. 1).
 FT SEQUENCE 319 AA; 36965 MW; 2DE880E9BFE19BC9 CRC64;
 SQ
 Query Match 51.5%; Score 35; DB 1; Length 319;
 Best Local Similarity 54.5%; Pred. No. 93;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AGLICPDPRP 11
 Db 58 SNLRCPDHPHPL 68
 RESULT 42
 ID ETFA_CLOTS STANDARD; PRT: 330 AA.
 AC E11N53; 1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE ELECTRON TRANSFER FLAVOPROTEIN ALPHA-SUBUNIT (ALPHA-ETP) (ELECTRON
 DE TRANSFER FLAVOPROTEIN LARGE SUBUNIT) (ETPLS).
 GN ETFA.
 OS Clostridium thermosaccharolyticum (Thermoanaerobacterium
 OS Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Thermoanaerobacter group; Thermoanaerobacterium.
 NCBI_TaxID=1517;
 RN
 SEQUENCE FROM N.A.
 RC STRANDSM 571; Brannenmeier K., Staudenbauer W.L.;
 RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RL -!- FUNCTION: THE ELECTRON TRANSFER FLAVOPROTEIN SERVES AS A SPECIFIC
 CC ELECTRON ACCEPTOR FOR OTHER DEHYDROGENASES. IT TRANSFERS THE
 CC ELECTRONS TO THE MAIN RESPIRATORY CHAIN VIA ETP-UBIQUINONE
 CC OXIDOREDUCTASE (ETP DEHYDROGENASE) (BY SIMILARITY).
 CC -!- SUBSTRATE: NADPH AND NADH.
 CC -!- SIMILARITY: BELONGS TO THE ETP ALPHA-SUBUNIT / FIXB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z82038; CAB04791.1; -;
 DR EMBL; Z92374; CAB07498.1; -;
 DR HSSP; P13804; ETP.
 DR PIR; P13804; ETP.
 DR PFAM; PF00756; ETP_alpha; 1.
 DR PROSITE; PS00696; ETP_ALPHA; 1.

KW Electron transport; Flavoprotein; FAD.
 FT NP_BIND 270 298 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 330 AA; 35598 MW; CB2495B44EFA43C CRC64;
 Query Match 51.5%; Score 35; DB 1; Length 330;
 Best Local Similarity 60.0%; Pred. No. 96;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AGLICPDPRP 10
 Db 147 ATIKCPDKRP 156
 RESULT 43
 ID YT35_STRFR STANDARD; PRT: 348 AA.
 AC P20186;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 34, Last sequence update)
 DE HYPOPHYSICAL 35.5 KDA PROTEIN IN TRANSPOSON TN4556.
 OS Streptomyces fradiae.
 OC Bacteria; Firmicutes; Actinobacteriae; Actinobacteridae;
 OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
 NCBI_TaxID=1906;
 RN
 SEQUENCE FROM N.A.
 RC TRANSPOSON-TN4556;
 RX Siemieniuk D.R., Slightom J.L., Chung S.T.;
 RA "Nucleotide sequence of Streptomyces fradiae transposable element
 RT TN4556: a class II transposon related to tn3."
 RL Gene 86:11-15(1990).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M29297; AA88561.1; -;
 DR PIR; JQ0431; JQ0431.
 KW Hypothetical protein; Transposable element.
 SQ SEQUENCE 348 AA; 35520 MW; 3bdf5d83abb8b92a CRC64;
 Query Match 51.5%; Score 35; DB 1; Length 348;
 Best Local Similarity 70.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 AGLICPDPRP 10
 Db 228 AGSLLPAPRP 237
 RESULT 44
 ID UL33_RCMVM STANDARD; PRT: 387 AA.
 AC UL33_RCMVM
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE G-PROTEIN COUPLED RECEPTOR HOMOLOG R33.
 DE Rat cytomegalovirus (strain Mastricht).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Cytomegalovirinae; Murineglovirus.
 NCBI_TaxID=79700;
 RN
 SEQUENCE FROM N.A.
 RP Beisser P.B., Vink C., Bruggeman C.A.;

```
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF232689; AAC58815.1; -
DR GCRdb; GCR_1283; -
DR InterPro; IPR000276; -
DR Pfam; PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Transmembrane.
FT DOMAIN 1 33 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 34 61 1 (POTENTIAL).
FT DOMAIN 62 71 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 72 94 2 (POTENTIAL).
FT DOMAIN 95 107 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 108 129 3 (POTENTIAL).
FT DOMAIN 130 150 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 151 169 4 (POTENTIAL).
FT DOMAIN 170 204 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 205 224 5 (POTENTIAL).
FT DOMAIN 225 244 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 245 268 6 (POTENTIAL).
FT DOMAIN 269 285 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 286 309 7 (POTENTIAL).
FT DOMAIN 310 387 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 20 20 POLY-PRO.
FT SEQUENCE 387 AA; 43159 MW; EF80D3F10344D6AE CRC64;
Query Match 51.5%; Score 35; DB 1; Length 387;
Best Local Similarity 54.5%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 AGLLCPDPRPL 11
| 1 1 | | |
Db 159 ASLMCASPAPL 169
-----
RESULT 45
ID GLNA_HALVO STANDARD; PRT; 454 AA.
AC P43386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLUTAMINE SYNTHETASE (EC 6.3.1.2) (GLUTAMATE--AMMONIA LIGASE) (GS).
GN GLNA.
OS Halobacterium volcanii (Haloferax volcanii).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloferax.
OX NCBI_TaxID=2246;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DS2;
RX MEDLINE=94365840; PubMed=7916055;
RA Brown J.R., Masuchi Y., Robb F.T., Doolittle W.F.;
RT "Evolutionary relationships of bacterial and archaeal glutamine
RT synthetase genes.";
RL J. Mol. Evol. 38:566-576(1994).
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + NH(3) = ADP + GLUTAMINE +
CC ORTHOPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03029; AAC43489.1; -
DR HSSP; P06201; 2LGS.
DR InterPro; IPR001691; -
DR Pfam; PF00120; gln-synt; 1.
DR PROSITE; PS00180; GLNA_1; 1.
DR PROSITE; PS00181; GLNA_ATP; 1.
KW Ligase.
FT BINDING 384 384 AMP (UNDER CONDITIONS OF ABUNDANT
FT GLUTAMINE) (BY SIMILARITY).
FT SEQUENCE 454 AA; 50696 MW; F07ECC3133B07722 CRC64;
Query Match 51.5%; Score 35; DB 1; Length 454;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GLLCDDP 8
| | | | |
Db 372 GLDCDDP 378
-----
Search completed: June 28, 2001, 11:57:25
Job time: 266 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:55:30 ; Search time 72.61 Seconds
(without alignments)

Title: US-09-439-313-558
Perfect score: 68
Sequence: 1 AGLLCPPDRPLE 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters:

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

```
Database : STREMBL_16:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.spore:*
12: sp.unclassified:*
13: sp.vertebrate:*
14: sp.virus:*
```

SUMMARIES

Result No.	Score	% Match	Query Length	DB Length	ID	Description
1	46	67.6	782	4	Q9UID3	Q9uid3 homo sapien
2	44	64.7	205	4	Q9Y4U5	Q9y4u5 homo sapien
3	44	64.7	757	5	Q22574	Q22574 caenorhabdi
4	43	63.2	362	13	Q92074	Q92074 g beta-1.4-
5	41	60.3	271	4	Q9HBS0	Q9hbs0 homo sapien
6	41	60.3	367	2	P95188	P95188 mycobacteri
7	41	60.3	392	2	O53870	O53870 mycobacteri
8	41	60.3	362	2	O53854	O53854 mycobacteri
9	41	60.3	568	14	Q69321	Q69321 marek's dis
10	41	60.3	574	6	Q62759	Q62759 sus scrofa
11	41	60.3	711	11	Q9ESJ4	Q9esj4 mus musculu
12	41	60.3	729	4	Q9UJ93	Q9uj93 homo sapien
13	41	60.3	1952	4	Q9UJ92	Q9uj92 homo sapien
14	41	60.3	2135	4	Q43157	Q43157 homo sapien
15	41	60.3	2135	4	Q9UIV7	Q9uiv7 homo sapien
16	40	58.8	140	2	Q915G7	Q915g7 pseudomonas
17	40	58.8	210	10	Q9S1Y1	Q9s1y1 arabidopsis
18	40	58.8	261	2	O86349	O86349 mycobacteri
19	40	58.8	328	14	Q68387	Q68387 human cytOm

ALIGNMENTS

RESULT	1
Q9UID3	PRELIMINARY;
ID	Q9UID3
AC	PRT; 782 AA.
DT	01-MAY-2000 (TREMBlrel. 13, Created)
DD	01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE	01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE	ANG2.
DN	GN ANG2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae
NCBI_TaxID=9606;	[1].
RN	SEQUENCE FROM N.A.
RP	MEDLINE=98277456; PubMed=9615229;
RX	Lemmens I.H., Kas K., Merregaert J., Van de Ven W.J.
RA	"Identification and molecular characterization of T
RT	gene cluster on human chromosome liq13.";
RL	Genomics 49:437-442(1998).
RF	[2]
RP	SEQUENCE FROM N.A.

DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DE HYPOTHETICAL 29.6 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
 RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
 RA Yu J., Han L.H.;
 RT "Novel Human cDNA clones with function of inhibiting cancer cell
 growth.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF217981; RAG17224.1;
 KW Hypothetical protein.
 SQ SEQUENCE 271 AA; 29630 MW; 31158B0F8F03B41F CRC64;

Query Match 60.3%; Score 41; DB 4; Length 271;
 Best Local Similarity 72.7%; Pred. No. 30;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 GLICPPRPLE 12
 I I I I I I I I
 DB 108 GQLCWSRPLE 118

RESULT 6
 P95188 PRELIMINARY; PRT; 362 AA.
 AC P95188;
 DT 01-MAY-1997 (TRENBLrel. 03, Created)
 DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
 DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
 DE PFLA
 GN PFLA OR RV3138 OR MTCY03A2.20C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 DR EMBL: Z83867; CAB06292.1;
 DR TuberculList; RV3138;
 SQ SEQUENCE 362 AA; 40432 MW; D5A3A234834D99EE CRC64;

Query Match 60.3%; Score 41; DB 2; Length 362;
 Best Local Similarity 66.7%; Pred. No. 39;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGLICPPR 9

Db 162 AGYNCPEPR 170
 I I I I I I I I

RESULT 7
 O53870 PRELIMINARY; PRT; 397 AA.
 AC O53870;
 DT 01-JUN-1998 (TRENBLrel. 06, Created)
 DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
 DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)
 DE PUTATIVE AMINOTRANSFERASE.
 GN RV0858C OR MTV043.51C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 DR EMBL: AL022004; CAAL7864.1;
 DR HSSP; Q56232; IBJW.
 DR TuberculList; RV0858C;
 DR InterPro; IPR001511;
 DR Pfam; PF00155; aminotran.1; 1.
 KW Transferase; Aminotransferase.
 SQ SEQUENCE 397 AA; 42209 MW; 2667A0F603D1C4F5 CRC64;

Query Match 60.3%; Score 41; DB 2; Length 397;
 Best Local Similarity 87.5%; Pred. No. 42;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LCPDPRPL 11
 I I I I I I I I
 DB 318 LCADPRPL 325

RESULT 8
 O53854 PRELIMINARY; PRT; 430 AA.
 AC O53854;
 DT 01-JUN-1998 (TRENBLrel. 06, Created)
 DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
 DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
 DE PUTATIVE MEMBRANE PROTEIN.
 GN RV0842 OR MTV043.35.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 DR EMBL: Z83867; CAB06292.1;
 DR TuberculList; RV3138;
 SQ SEQUENCE 362 AA; 40432 MW; D5A3A234834D99EE CRC64;

RA Rutter S., Seeger K., Skelton S., Squares S., Scares R., Sulston J.E.,
 RA K. K., Whittaker J., Whittaker J., Whittaker J., Whittaker J.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 DR EMBL: AL022004; CAAL17648.1;
 DR TubercuList; RV0842;
 SQ SEQUENCE 430 AA; 43854 MW; B904FA823AD2F67F CRC64;

Query Match 60.3%; Score 41; DB 2; Length 430;
 Best Local Similarity 63.6%; Pred. NO. 45;
 Matches 7; Conservative 2; Mismatches 0; Gaps 0;

OY 1 AGLICPDPRPL 11
 Db 120 AGLICPDPRV 130

RESULT 9
 Q69321 PRELIMINARY; PRT; 568 AA.
 AC Q69321;
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
 DE UL46H (UL46H TEGMENT PHOSPHOPROTEIN-LIKE PROTEIN).
 GN UL46H OR MDV059. herpesvirus (strain GA) (MDHV), and
 GN Turkey herpesvirus.
 OS Turkey herpesvirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicelloviruses.
 NCBI_TaxID=10388, 10390;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN-GA; Turkey herpesvirus; STRAIN-GA;
 RC MEDLINE-9339438; PubMed=8397281;
 RA Yanagida N., Yoshida S., Nazerian K., Lee L.F.;
 RT "Nucleotide and predicted amino acid sequences of Marek's disease
 RT virus homologues of herpes simplex virus major tegument proteins."
 RL J. Gen. Virol. 74:1837-1845(1993).
 [2]
 SEQUENCE FROM N.A.
 RC STRAIN-GA; Turkey herpesvirus; STRAIN-GA;
 RC MEDLINE-92237304; PubMed=1315048;
 RA Jones D., Lee L., Liu J.L., Kung H.J., Tillotson J.K.;
 RT "Marek disease virus encodes a basic-leucine zipper gene resembling
 RT the fos/jun oncogenes that is highly expressed in lymphoblastoid
 RT tumors."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046(1992).
 [3]
 SEQUENCE FROM N.A.
 RC STRAIN-GA; Turkey herpesvirus; STRAIN-GA;
 RC SPECIES-Turkey herpesvirus; STRAIN-GA;
 RA Lee L.F., Wu P., Sul D., Ren D., Kung H.J.;
 RT "The Complete UL Sequence of Serotype I Marek's Disease Virus."
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2000).
 [4]
 SEQUENCE FROM N.A.
 RC STRAIN-GA; Turkey herpesvirus; STRAIN-GA;
 RC MEDLINE-20392152; PubMed=10933706;
 RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
 RT "The genome of a very virulent Marek's disease virus."
 RL J. Virol. 74:7980-7988(2000).
 [5]
 SEQUENCE FROM N.A.
 RC STRAIN-GA; Turkey herpesvirus; STRAIN-GA;
 RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Rock D.L., Kutish G.F.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: L10283; AAA03151.1;
 DR EMBL: AF147806; AAF66781.1;
 DR EMBL: AF243438; AAG14239.1;
 SQ SEQUENCE 568 AA; 63924 MW; 6D88474526703B58 CRC64;

Query Match 60.3%; Score 41; DB 14; Length 568;
 Best Local Similarity 54.5%; Pred. NO. 58;
 Matches 6; Conservative 3; Mismatches 0; Gaps 0;

OY 2 GLICPDPRPLE 12
 Db 479 GLICPDPRPAD 489

RESULT 10
 O62759 PRELIMINARY; PRT; 574 AA.
 AC O62759;
 DT 01-AUG-1998 (TREMREL. 07, Created)
 DT 01-AUG-1998 (TREMREL. 07, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
 DE EPIDERMAL GROWTH FACTOR PRECURSOR (FRAGMENT).
 GN EGF.
 OS Sus scrofa (Pig).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 NCBI_TaxID=9823;
 [1]
 SEQUENCE FROM N.A.
 RC STRAIN-DANISH LANDRACE;
 RA Joergensen P.E., Jensen L.G., Sorensen B.S., Poulsen S.S., Nexoe E.;
 RT "Cloning and characterization of the porcine epidermal growth factor
 RT EMBL: AF053364; AAC14024.1; 0:0-0(1998).
 DR HSP; P07204; IFGD.
 DR InterPro; IPR000033;
 DR InterPro; IPR000152;
 DR InterPro; IPR000152;
 DR InterPro; IPR000152;
 DR Pfam; PF00058; ldl_recept_b; 4.
 DR PROSITE; PS001010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE; PS01186; EGF_2; 4.
 DR PROSITE; PS01187; EGF_CA; 1.
 DR SMART; SM00001; EGF_Like; 1.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Repeat.
 FT NON_TER
 FT NON_TER
 FT NON_TER
 SQ SEQUENCE 574 AA; 62627 MW; 4710A4E2E25167D CRC64;

Query Match 60.3%; Score 41; DB 6; Length 574;
 Best Local Similarity 66.7%; Pred. NO. 59;
 Matches 6; Conservative 1; Mismatches 2; Indels 0;

OY 2 GLICPDPRP 10
 Db 560 GRICPDPRP 568

RESULT 11
 Q9ESJ4 PRELIMINARY; PRT; 711 AA.
 AC Q9ESJ4;
 DT 01-MAR-2001 (TREMREL. 16, Created)
 DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
 DE N-WASP BINDING PROTEIN WISH.
 GN WISH.
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 RC "Kurooka M., Saito S., Miki H., Fukami K., Endo T., Takenawa T.;
 RA "Wish, a novel protein, induces ATP2/3 complex
 RT activation independent of Cdc42."
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF130313; AAF36503.1; -;
SQ SEQUENCE 711 AA; 78706 MW; FE700EA466A251BD CRC64;

Query Match 60.3%; Score 41; DB 11; Length 711;
Best Local Similarity 77.8%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPPDRPL 11
Db 246 LLCPSPL 254

RESULT 12

Q9UJ93 ID Q9UJ93 PRELIMINARY; PRT; 729 AA.
AC Q9UJ93;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PLEXIN-B1/SEP RECEPTOR PRECURSOR.
GN PLEXIN-B1/SEP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EPITHELIUM;
RX MEDLINE=99449305; PubMed=10520995;
RA Tamagnone L., Artigiani S., Chen H., He Z., Ming G.L., Song H.L.,
RA Chedotal A., Winberg M.L., Goodman C.S., Poo M.M., Tessier-Lavigne M.,
RA Comoglio P.M.;
RA "Plexins are a large family of receptors for transmembrane, secreted
RT and GPI-anchored semaphorins in vertebrates.";
RL Cell 99:71-80(1999).
DR EMBL; AJ011415; CAB56221.1; -;
DR InterPro; IPR001627; -;
DR Pfam; PF01403; Sema; 2.
DR Pfam; PF01437; Plexin_repeat; 1.
DR SMART; SM00423; PSI; 1.
KW Signal; Receptor.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 729 PLEXIN-B1/SEP RECEPTOR.
SQ SEQUENCE 729 AA; 78394 MW; 4AE56E0DBAE6D75 CRC64;

Query Match 60.3%; Score 41; DB 4; Length 729;
Best Local Similarity 50.0%; Pred. No. 73;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLCCDPRPLE 12
Db 584 SGWCFSPDPSE 595

RESULT 13

Q9UJ92 ID Q9UJ92 PRELIMINARY; PRT; 1952 AA.
AC Q9UJ92;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE PLEXIN-B1/SEP RECEPTOR PRECURSOR.
GN PLEXIN-B1/SEP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99449305; PubMed=10520995;

RA Tamagnone L., Artigiani S., Chen H., He Z., Ming G.L., Song H.L.,
RA Chedotal A., Winberg M.L., Goodman C.S., Poo M.M., Tessier-Lavigne M.,
RA Comoglio P.M.;
RA "Plexins are a large family of receptors for transmembrane, secreted
RT and GPI-anchored semaphorins in vertebrates.";
RL Cell 99:71-80(1999).
DR EMBL; AJ011415; CAB56222.1; -;
DR InterPro; IPR000566; -;
DR InterPro; IPR001627; -;
DR InterPro; IPR002165; -;
DR InterPro; IPR002589; -;
DR InterPro; IPR002909; -;
DR InterPro; IPR003006; -;
DR Pfam; PF01403; Sema; 2.
DR Pfam; PF01437; Plexin_repeat; 2.
DR Pfam; PF01833; TIG; 3.
DR ProDom; PD003489; -; 1.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR SMART; SM00423; PSI; 1.
KW Signal; Receptor.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 1952 PLEXIN-B1/SEP RECEPTOR.
SQ SEQUENCE 1952 AA; 214409 MW; FCAD0630E128EE9C CRC64;

Query Match 60.3%; Score 41; DB 4; Length 1952;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLCCDPRPLE 12
Db 584 SGWCFSPDPSE 595

RESULT 14

O43157 ID O43157 PRELIMINARY; PRT; 2135 AA.
AC O43157;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE KIA0407.
GN KIA0407.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA Ishikawa K., Nagase T., Nakajima D., Seki N., Ohira M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007867; BAA23703.1; -;
DR InterPro; IPR000566; -;
DR InterPro; IPR001627; -;
DR InterPro; IPR001936; -;
DR InterPro; IPR002165; -;
DR InterPro; IPR002909; -;
DR InterPro; IPR003006; -;
DR Pfam; PF01403; Sema; 2.
DR Pfam; PF01437; Plexin_repeat; 2.
DR Pfam; PF01833; TIG; 3.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR SMART; SM00429; IPT; 1.
SQ SEQUENCE 2135 AA; 232295 MW; 12A81B68AF1D340F CRC64;

Query Match 60.3%; Score 41; DB 4; Length 2135;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;


```
Best Local Similarity 66.7%; Pred. No. 35;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLCPPDRPL 11
Db 169 LVCPPDRSI 177

RESULT 18
ID O86349 PRELIMINARY; PRT; 261 AA.
AC O86349;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 27.4 KDA PROTEIN.
GN RV1457C OR MTV007.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Sultston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Parkhill J.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021184; CAA15987.1; -.
DR TuberculList; RV1457C; -.
DR InterPro; IPR000005; -.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 261 AA; 27370 MW; FF2074B7B862C021 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 261;
Best Local Similarity 70.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGLCCPDRP 10
Db 11 AGTFSPDRP 20

RESULT 19
ID Q68387 PRELIMINARY; PRT; 328 AA.
AC Q68387;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ORF UL135.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOLEDO;

RX MEDLINE=96099416; PubMed=8523595;
RA Cha T.A., Tom E., Kemble G.W., Duke G.M., Mocarski E.S., Spaete R.R.;
RT "Human cytomegalovirus clinical isolates carry at least 19 genes not
RT found in laboratory strains.";
RL J. Virol. 70:78-83(1996).
DR EMBL; U33331; AAA85874.1; -.
SQ SEQUENCE 328 AA; 35735 MW; E5FBF20F5E6841A CRC64;

Query Match 58.8%; Score 40; DB 14; Length 328;
Best Local Similarity 77.8%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GLCCPDRP 10
Db 273 GLSCPCPRP 281

RESULT 20
ID Q9L1A1 PRELIMINARY; PRT; 355 AA.
AC Q9L1A1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE (FRAGMENT).
GN SC1068.01C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL158057; CAB76275.1; -.
FT NON_TER 355 355
SQ SEQUENCE 355 AA; 38866 MW; 93680C4EDB499052 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 355;
Best Local Similarity 87.5%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLCPPDRP 10
Db 235 LLCVDRP 242

RESULT 21
ID Q9RNX2 PRELIMINARY; PRT; 522 AA.
AC Q9RNX2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PUTATIVE EFFLUX PUMP COMPONENT MTRF.
GN MTRF.
```

OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FA19;
 RA veal W.L., Shafer W.M.; required for high-level antimicrobial
 RT novel protein (MTRF) to the EMBL/GenBank/DDBJ databases.
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
 DR EMBL: AF176820; AAD51362.2; --
 DR InterPro: IPR001589; --
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1;
 DR SEQUENCE 522 AA; 56131 MW; 1351EC9A32E6AA43 CRC64;
 SO SEQUENCE 522 AA; 56131 MW; 1351EC9A32E6AA43 CRC64;
 Query Match 58.8%; Score 40; DB 2; Length 522;
 Best Local Similarity 70.0%; Pred. No. 79;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 GLCCPDPRL 11
 DB 48 GLSVDPDPV 57
 RESULT 22
 Q9JY63 PRELIMINARY; PRT; 522 AA.
 ID O9JY63
 AC O9JY63
 DT 01-MAY-2000 (TREMREL. 13, Created)
 DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
 DT 01-JUN-2000 (TREMREL. 14, Last annotation update)
 DE PUTATIVE EFFLUX PUMP COMPONENT MTRF.
 GN MTRF.
 GS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EU75;
 RA veal W.L., Shafer W.M.; required for high-level antimicrobial
 RT resistance by the gonococcal mtr efflux pump;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
 DR EMBL: AF176821; AAD51363.2; --
 DR InterPro: IPR001589; --
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1;
 DR SEQUENCE 522 AA; 56175 MW; 7A7BF0BEF128F7E9 CRC64;
 SO SEQUENCE 522 AA; 56175 MW; 7A7BF0BEF128F7E9 CRC64;
 Query Match 58.8%; Score 40; DB 2; Length 522;
 Best Local Similarity 70.0%; Pred. No. 79;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 GLCCPDPRL 11
 DB 48 GLSVDPDPV 57
 RESULT 23
 Q9JY63 PRELIMINARY; PRT; 522 AA.
 ID O9JY63
 AC O9JY63
 DT 01-OCT-2000 (TREMREL. 15, Created)
 DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
 DE EFFLUX PUMP COMPONENT MTRF.
 GN MTRF.
 GS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=4391;
 RA J.L.

RC STRAIN=WC58 / SEROGROUP B;
 MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Citroner H., Clark E.B.,
 RA Gillis M., Scarlata V., Maignani V., Pizzano M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58";
 RL Science 287:1809-1815(2000).
 DR EMBL: AE002522; AAF62337.1; --
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1;
 DR SEQUENCE 522 AA; 56240 MW; 41CCB4AC20E190D1 CRC64;
 SO SEQUENCE 522 AA; 56240 MW; 41CCB4AC20E190D1 CRC64;
 Query Match 58.8%; Score 40; DB 2; Length 522;
 Best Local Similarity 70.0%; Pred. No. 79;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 GLCCPDPRL 11
 DB 48 GLSVDPDPV 57
 RESULT 24
 Q9JY63 PRELIMINARY; PRT; 522 AA.
 ID O9JY63
 AC O9JY63
 DT 01-OCT-2000 (TREMREL. 15, Created)
 DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
 DE HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN.
 GN NMA1973.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
 MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis G., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
 RA Ragsdale M., Rastall M., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis Z2491";
 RL Nature 404:502-506(2000).
 DR EMBL: AL162757; CAB85193.1; --
 DR PROSITE: PS00019; ACTININ_1; UNKNOWN_1;
 DR SEQUENCE 522 AA; 56139 MW; 412A9DF454BE94DC CRC64;
 SO SEQUENCE 522 AA; 56139 MW; 412A9DF454BE94DC CRC64;
 Query Match 58.8%; Score 40; DB 2; Length 522;
 Best Local Similarity 70.0%; Pred. No. 79;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 2 GLCCPDPRL 11
 DB 48 GLSVDPDPV 57
 RESULT 25
 Q9JY63 PRELIMINARY; PRT; 528 AA.
 ID O49602
 AC O49602

```

DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE IMPORTIN ALPHA-LIKE PROTEIN (FRAGMENT).
GN IMPA-4.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=VEGETATIVE TISSUE;
RA Schledz M., Leclerc D., Neuhaus G., Merkle T.;
RL Plant Physiol. 116:868-868(1998).
DR EMBL; Y14616; CAA74966.1; -.
DR HSSP; Q02821; IBK5.
DR Mendel; 24707; Arath; 2729; 24707.
DR InterPro; IPR000225; -.
DR InterPro; IPR002652; -.
DR InterPro; IPR003006; -.
DR Pfam; PF00514; Armadillo_seg; 8.
DR Pfam; PF01749; IBB; 1.
DR PROSITE; PS01176; ARM_REPEAT; 2.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR SMART; SM00185; ARM; 1.
FT NON_TER 1
FT 1
SQ SEQUENCE 528 AA; 58265 MW; D794CAICE959E731 CRC64;

Query Match 58.8%; Score 40; DB 10; Length 528;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LICPDPR 9
Db 418 LICPDPR 424

RESULT 26
ID O80480 PRELIMINARY; PRT; 538 AA.
AC O80480;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE T12M4.2 PROTEIN.
GN T12M4.2.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Toriumi M., Yu G., Oji, O, Kwan, A,
RA Liu S., Li J., Araujo R., Au M., Brendel V., Buehler E., Conway A.B.,
RA Conway A.R., Dewar K., Feng J., Kim C., Kurtz D., Li Y., Palm C.J.,
RA Shinn P., Sun H., Davis R.W., Ecker J.R., Federspiel N.A.,
RA Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC T12M4 sequence, complete
sequence.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;

```

```

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC003114; AAC24079.1; -.
DR HSSP; Q02821; IBK5.
DR Mendel; 31141; Arath; 2729; 31141.
DR InterPro; IPR000225; -.
DR InterPro; IPR002652; -.
DR Pfam; PF00514; Armadillo_seg; 8.
DR Pfam; PF01749; IBB; 1.
DR PROSITE; PS01176; ARM_REPEAT; 3.
DR SMART; SM00185; ARM; 1.
SQ SEQUENCE 538 AA; 59445 MW; C342198841B87CCA CRC64;

Query Match 58.8%; Score 40; DB 10; Length 538;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LICPDPR 9
Db 428 LICPDPR 434

RESULT 27
ID Q9RRT1 PRELIMINARY; PRT; 633 AA.
AC Q9RRT1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE ABC TRANSPORTER, ATP-BINDING PROTEIN, MSBA FAMILY.
GN DR2404.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE002071; AAF11948.1; -.
DR HSSP; P13569; INBD.
DR TIGR; DR2404; -.
DR InterPro; IPR001687; -.
DR InterPro; IPR003439; -.
DR InterPro; IPR003593; -.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
KW ATP-binding.
SQ SEQUENCE 633 AA; 69337 MW; DB8549554C15EB32 CRC64;

Query Match 58.8%; Score 40; DB 2; Length 633;
Best Local Similarity 87.5%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LICPDPR 11
Db 369 LAPDPR 376

```

```

RESULT 28
Q9U122 PRELIMINARY; PRT; 1301 AA.
AC Q9U122;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAY-2001 (TrEMBLrel. 13, Last annotation update)
DE HYPOPHETICAL 137.3 KDA PROTEIN.
GN L4768.04
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
ON NCBI_TaxID=5664;
RX [1]
RC SEQUENCE FROM N.A.
RC STRAIN=FRIDLIN.
RA Bothe G., Pohl T., Ivens A.C., Lawson D., Murphy L., Quail M.,
RA Rajadrem M.A.C., Barrell B.G.;
RA Submitted (MDC-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RP STRAIN=FRIDLIN.
RX [1]
RC SEQUENCE FROM N.A.
RC STRAIN=FRIDLIN.
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL EMBL; AL132763; CAB59861.1; -.
DR Enterp000226004752ain. 1.
DR PRINTS; PR00380; KINESIN_MOTOR_DOMAIN2; 1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
DR SMART; SM00129; Kisc; 1.
DR ATP-binding; Coiled coil; Hypothetical protein; Microtubules;
KW Motor protein.
KW SEQUENCE 1301 AA; 137311 MW; 6822477834A0B521 CRC64;
SQ

```

Query Match 58.8%; Score 40; DB 5; Length 1301;
 Best Local Similarity 80.0%; Pred. NO. 1.8e+02;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 LLCPPRPLE 12
 DB 109 LLVAPRPLE 118

```

RESULT 29
Q9437 PRELIMINARY; PRT; 1624 AA.
AC Q9437;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CG4320 PROTEIN.
GN CG4320hilla melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
ON NCBI_TaxID=7227;
RX [1]
RC SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman V.R., Yandell M.D., Zhang Q., Chen B.D.,
RA Watson K.R., Rogers J.C., Gish W.B., Granger D.A., Pfeiffer B.D.,
RA Wang K.H., Doyle C.B., Baxter E.G., Helgert G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

```

```

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P., M.,
RA de Pablos B., Delcher A., Deng Z., Dey S., Dey S., Dey S., Dey S.,
RA Dodson K., Doolittle R., Doolittle R., Doolittle R., Doolittle R.,
RA Doolittle R., Doolittle R., Doolittle R., Doolittle R., Doolittle R.,
RA Poser C.J., Gabrielian A.E., Garq N.S., Gelbart W.M., Glessner K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hatalin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Hatalin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Hatalin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Kimmel B., Kalush F.C., Karpen G.H., Kaur A., Kaur A., Kaur A.,
RA Kimmel B., Kalush F.C., Karpen G.H., Kaur A., Kaur A., Kaur A.,
RA Kimmel B., Kalush F.C., Karpen G.H., Kaur A., Kaur A., Kaur A.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount D.R., Nelson K.A., Nixon K., Nusker D.R., Pacier J.M., D.L.,
RA Palazzolo M., Pittman G.S., Pan S., Pan S., Pan S., Pan S.,
RA Reiner B.C., Sidelmann I., Simpson M., Skupski M.P., Smith T.,
RA Sier E.C., Sidelmann I., Simpson M., Skupski M.P., Smith T.,
RA Sier E.C., Sidelmann I., Simpson M., Skupski M.P., Smith T.,
RA Sier E.C., Sidelmann I., Simpson M., Skupski M.P., Smith T.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zhao L.,
RA Cheng X.H., Hong E.W., Rubi G.M., Venter J.C., Smith H.O.,
RT "The genome sequence of Drosophila melanogaster."
RL EMBL; AF003436; AAF46122.1; -.
DR Flybase; FBgn0029840; CG4320.
DR InterPro; IPR001680; 5.
DR SMART; SM00320; WMD40; 1.
KW Repeat; WD repeat.
KW SEQUENCE 1624 AA; 177525 MW; B7B06FD9E830739C CRC64;
SQ

```

Query Match 58.8%; Score 40; DB 5; Length 1624;
 Best Local Similarity 87.7%; Pred. NO. 2.2e+02;
 Matches 0; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 LLCPPRPL 11
 DB 876 VMCPDRDL 884

```

RESULT 30
P93810 PRELIMINARY; PRT; 399 AA.
ID P93810;
AC P93810;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DE 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOPHETICAL 44.7 KDA PROTEIN.
GN F19P19.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
ON NCBI_TaxID=3702;
RX [1]
RC SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Osborne B.I., Toriumi M., Yu G., Oji O., Buehler E.,
RA Conway A.B., Conway A.R., Dewar K., Teng J., Kim C.,
RA Chinn N., Sun H., Davis R.W., Ecker J.R., Federici N.A.,
RA Chinn N., Sun H., Davis R.W., Ecker J.R., Federici N.A.,
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RC

```

RA Theologis A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC000104; AAB70424.1; -
KW Mendel; 12684; Arath;1884;12684.
SQ SEQUENCE 399 AA; 44689 MW; 680ED6D4BA3ED347 CRC64;

Query Match 58.1%; Score 39.5; DB 10; Length 399;

Best Local Similarity 63.6%; Pred. No. 75;

Matches 7; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 2 GLLCPDP-RPL 11

|| ||||| :|

Db 252 GLFCDPPIKPI 262

RESULT 31

ID Q63552

AC Q63552 PRELIMINARY; PRT; 146 AA.

DT 01-NOV-1996 (TRENBLrel. 01, Created)

DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TRENBLrel. 08, Last annotation update)

DE SMR1-ALPHA1.

GN VCS-ALPHA1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_TaxID=10116;

RN SEQUENCE FROM N.A.

RC TISSUE-SUBMANDIBULAR GLAND;

RA MEDLINE=95169272; PubMed=8754212;

RA Singer M., Courtney Y., Rougeon F.;

RT "Recent evolution of genes encoding the prohormone-like protein SMR1

in the rat submandibular gland.";

RL DNA Cell Biol. 14:137-144(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE-SUBMANDIBULAR GLAND;

RA Courtney Y., Singer M., Rosinski-Chupin I., Rougeon F.;

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE-SUBMANDIBULAR GLAND;

RA MEDLINE=96276306; PubMed=8754212;

RA Courtney Y., Singer M., Rosinski-Chupin I., Rougeon F.;

RT "Episodic evolution and rapid divergence of members of the rat

multigene family encoding the salivary prohormone-like protein SMR1.";

RL Moll. Biol. Evol. 13:758-766(1996).

DR EMBL; X77819; CAA54834.1; -

FT CHAIN 23 146 SMR1-ALPHA1.

SQ SEQUENCE 146 AA; 16282 MW; BCBC5E360A1C50E8 CRC64;

Query Match 57.4%; Score 39; DB 11; Length 146;

Best Local Similarity 77.8%; Pred. No. 37;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 LLCPPDPRL 11

| |||||

Db 98 LTAPDPRL 106

RESULT 32

ID Q9V3K5

AC Q9V3K5 PRELIMINARY; PRT; 262 AA.

DT 01-MAY-2000 (TRENBLrel. 13, Created)

DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)

DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)

DE BG:DS06874.1 PROTEIN.

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003645; AAF53408.1; -

GN BG:DS06874.1 OR CG4691.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Y, AND CN BW SP;

RA MEDLINE=99403001; PubMed=10471707;

RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazej R., Davis T.,

RA Doyle C., Galle R., George R.A., Harris N.L., Hoskins R.A.,

RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,

RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitelaw K.,

RA Celniker S., Rubin G.M.;

RA Drosophila melanogaster: the sequence of a 2.9-Mb region of the genome of

the Adh region.";

RL Genetics 153:179-219(1999).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Y, AND CN BW SP;

RA Celniker S.E., Agbayani A., Arcaina T.T., Baxter E., Blazej R.G.,

RA Butenhoff C., Champe M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,

RA Farfan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,

RA Houston K.A., Hummasti S.R., Karra K., Kearney L., Kim E., Lee B.,

RA Lewis S., Li P., Lomotan M.A., Mazda P., Moshrefi A.R., Moshrefi M.,

RA Nixon K., Pacleb J.M., Park S., Pfeiffer B., Poon L., Sequeira A.,

RA Sethi H., Snir E., Svirska R.R., Wan K.H., Weinburg T., Zhang R.,

RA Zierian L.L., Rubin G.M.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE003645; AAF53408.1; -

DR InterPro: IPR002652; -
 DR AF005120; TRENDEL 08, Created)
 DR PF01749; IBB 1
 DR PROSITE; PS00176; ARM_REPEAT; 4.
 DR SMART; SM00185; ARM; 1.
 DR SEQUENCE 532 AA; 58670 MW; DA80EC17E1ACF70E CRC64;
 SQ
 Query Match 57.4%; Score 39; DB 10; Length 532;
 Best Local Similarity 85.7%; Pred No. 1.2e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 3 LLCPPDR 9
 Db 421 LVCPPDR 427
 RESULT 38
 ID O81520 PRELIMINARY; PRT; 532 AA.
 AC O81520; 1998 (TRENDEL 08, Created)
 DT 01-NOV-1998 (TRENDEL 08, Last sequence update)
 DE IMPORTIN ALPHA.
 GN F3822.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Eudicotyledons; Core eudicotyledons; Rosidae; eurosids 1;
 OC Rosidae; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLOMBIA;
 RA MEDLINE=96373199; PubMed=8776900; S. Raikhel N.V.;
 RA "Nuclear import in permeabilized protoplasts from higher plants has
 RT unique features.";
 RL Plant Cell 8:1337-1352(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLOMBIA;
 RA Smith H.M. Hicks G.R. Raikhel N.V.;
 RA "Importin alpha from Arabidopsis thaliana is a nuclear import receptor
 RT that recognizes three classes of import signals.";
 RL Plant Physiol. 114:411-417(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLOMBIA;
 RA Raikhel N.V. Smith H.M.S.;
 RA Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLOMBIA;
 RA Smith H.M. Hicks G.R. Raikhel N.V. Wu D.;
 RA "Arabidopsis thaliana chromosome III BAC F3E22 genomic sequence.";
 RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF077528; AAC27644.1; -
 DR EMBL; AC023912; AAF63826.1; -
 DR M58613; 32157; IBK5th; 2729; 32157.
 DR InterPro; IPR002652; -
 DR PF01749; IBB 1.
 DR PROSITE; PS00176; ARM_REPEAT; 4.
 DR SMART; SM00185; ARM; 1.
 DR SEQUENCE 532 AA; 58644 MW; 9425EACB98D585C7 CRC64;
 SQ

DR InterPro: IPR002652; -
 DR AF005120; TRENDEL 08, Created)
 DR PF01749; IBB 1
 DR PROSITE; PS00176; ARM_REPEAT; 4.
 DR SMART; SM00185; ARM; 1.
 DR SEQUENCE 532 AA; 58670 MW; DA80EC17E1ACF70E CRC64;
 SQ
 Query Match 57.4%; Score 39; DB 10; Length 532;
 Best Local Similarity 85.7%; Pred No. 1.2e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 3 LLCPPDR 9
 Db 421 LVCPPDR 427
 RESULT 39
 ID O9SLX0 PRELIMINARY; PRT; 534 AA.
 AC O9SLX0; 2000 (TRENDEL 13, Created)
 DT 01-MAY-2000 (TRENDEL 13, Last sequence update)
 DE IMPORTIN ALPHA 1B.
 GN F3822.14.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE; TISSUE=LEAF;
 RA Jiang C., Shoji K., Matsuki R., Inagaki N., Ban H., Iwasaki T.,
 RA Imamoto N., Yoneda Y., Yamamoto N.;
 RA Cloning and functional analysis of a novel importin alpha homologue
 RT from rice (Oryza sativa).
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN EMBL; AB024311; BAA88950.1; -
 DR HSSP; Q02821; IBK5.
 DR InterPro; IPR000225; -
 DR InterPro; IPR002652; -
 DR PF01749; IBB 1.
 DR PROSITE; PS00176; ARM_REPEAT; 3.
 DR SMART; SM00185; ARM; 1.
 DR SEQUENCE 534 AA; 58532 MW; 90C8A3F7ADEF1720 CRC64;
 SQ
 Query Match 57.4%; Score 39; DB 10; Length 534;
 Best Local Similarity 85.7%; Pred No. 1.2e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 3 LLCPPDR 9
 Db 427 LVCPPDR 433
 RESULT 40
 ID O49600 PRELIMINARY; PRT; 535 AA.
 AC O49600; 1998 (TRENDEL 06, Created)
 DT 01-JUN-1998 (TRENDEL 06, Last sequence update)
 DE IMPORTIN ALPHA-LIKE PROTEIN.
 GN IMPA-2.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Eudicotyledons; Core eudicotyledons; Rosidae; eurosids 1;
 OC Rosidae; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=VEGETATIVE TISSUE;
 RA Schledz M., Lecierc D., Neuhaus G., Merkle T.;
 RL Plant Physiol. 114:868-868(1998).
 DR HSSP; Q02821; IBK5.
 DR M58613; 32157; IBK5th; 2729; 32157.
 DR InterPro; IPR000225; -
 DR PF01749; IBB 1.
 DR PROSITE; PS00176; ARM_REPEAT; 4.
 DR SMART; SM00185; ARM; 1.
 DR SEQUENCE 535 AA; 58644 MW; 9425EACB98D585C7 CRC64;
 SQ

DR InterPro: IPR002652; -
DR Pfam: PF00514; Armadillo_seg; 8.
DR Pfam: PF01749; IBB; 1.
DR PROSITE: PS0176; ARM_REPEAT; 3.
DR SMART: SM00185; ARM; 1.
SQ SEQUENCE 535 AA; 58858 MW; 029745BF1DB5A5A9 CRC64;

Query Match 57.4%; Score 39; DB 10; Length 535;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LCPDPR 9
I:|||||

Db 426 LVCPPDR 432

RESULT 41

ID Q9P200 PRELIMINARY; PRT; 740 AA.
AC Q9P200;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE KIAA1528 PROTEIN (FRAGMENT).
GN KIAA1528.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20277482; PubMed-10819331;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human
RT genes.XVII.The complete sequences of 100 new cDNA clones from brain
RT which code for large proteins in vitro."
RL DNA Res. 7:143-150(2000).
DR EMBL; AB040961; BAA96052.1; -
DR InterPro: IPR001304; -
DR Pfam: PF00097; zf-G3HC4; 1.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 740 AA; 79828 MW; 9D8B622DFF9397D CRC64;

Query Match 57.4%; Score 39; DB 4; Length 740;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 CPDPRPL 11
||| |||||

Db 22 CPQPRPL 28

RESULT 42

ID P91658 PRELIMINARY; PRT; 974 AA.
AC P91658;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE FURROWED.
GN FW OR CG1500.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Leshko-Lindsay L., Corces V.G.;

RL Development 0:0-0(0).
DR EMBL; U70770; AAB36703.1; -
DR HSP; P02749; LOUB.
DR FlyBase; FBgn0001083; fw.
DR InterPro: IPR000436; -
DR InterPro: IPR001304; -
DR Pfam: PF00084; sushi; 10.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; UNKNOWN_1.
DR PROSITE: PS00441; C_TYPE_LECTIN_2; 1.
DR SMART; SM00032; CCP; 1.
SQ SEQUENCE 974 AA; 107557 MW; A406335FB726757 CRC64;

Query Match 57.4%; Score 39; DB 5; Length 974;
Best Local Similarity 71.4%; Pred. No. 2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 5 CPDPRPL 11
|||||:

Db 293 CPDPQPI 299

RESULT 43

Q9VYR4
ID Q9VYR4 PRELIMINARY; PRT; 1124 AA.
AC Q9VYR4;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE FW GENE PRODUCT.
GN FW OR CG1500.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacile J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A., Ye J., Yeh R., F. Zeng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR HBB; AF003487; AAF48125.1;
 DR HBB; AF003489; AAF48125.1;
 DR FlyBase; FB000083; fw.
 DR InterPro; IPR001304;
 DR InterPro; IPR001304;
 DR Pfam; PF00059; lectin_C; 1.
 DR Pfam; PF00084; sushi; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 DR PROSITE; PS00615; C-TYPE LECTIN_2; 1.
 DR SMART; SM00033; LCP; 1.
 SQ SEQUENCE 1124 AA; 123051 MW; COEB0528A415F34A CRC64;

Query Match 57.4%; Score 39; DB 5; Length 1124;
 Best Local Similarity 71.4%; Pred. No. 2.3e+02; Indels 0; Gaps 0;
 Matches 5; Conservative

OY 5 CPDPRL 11
 DB 475 CPDPQPI 481
 |||||

RESULT 44
 OY97C6 PRELIMINARY; PRT; 1529 AA.
 AC OY97C6;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE ESPRASE 0000 (TREMBLrel. 14, Last annotation update)
 OS Aspergillus terreus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=33178;
 RN 11 JENGE FROM N.A.
 RC STRAIN=ATCC20542.
 RA Kennedy J., Auclair K., Kennedy S.G., Park C., Vederas J.C.,
 RA Hutchinson C.R.;
 RT "Accessory Proteins Modulate Polyketide Synthase Activity During
 Lovastatin Biosynthesis.";
 RL Science 0:000(1999)
 DR InterPro; IPR001087;
 SQ SEQUENCE 1529 AA; 170505 MW; 456FB780A6B531A3 CRC64;

Query Match 57.4%; Score 39; DB 3; Length 1529;
 Best Local Similarity 77.8%; Pred. No. 3e+02; Indels 0; Gaps 0;
 Matches 7; Conservative

OY 3 LLCPPRPL 11
 DB 1097 LLCPPDPNLL 1105
 |||||

RESULT 45
 OY97C6 PRELIMINARY; PRT; 1734 AA.
 AC OY97C6;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE ESPRASE 0000 (TREMBLrel. 16, Last annotation update)
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN 11 JENGE FROM N.A.
 RP MEDLINE=20347919; PubMed=10748113;
 RX Wang Y., Sugita S., Sudhof T.C.;
 RT "The RIM/NIM Family of Neuronal C2 Domain Proteins. Interactions with
 Rab3 and a new class of Src homology 3 domain proteins.";
 DR Biol. Cell 83:255-263(1995).
 DR EMBL; AF199338; AAF81660.1;
 DR InterPro; IPR001452;
 DR InterPro; IPR001777;
 DR Pfam; PF00018; SH3; 3.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR PROSITE; PS00002; SH3; 3.
 DR SMART; SM00060; FNS; 1.
 FT NON-TER
 SQ SEQUENCE 1734 AA; 188524 MW; B771ECBE900A0AC7 CRC64;

Query Match 57.4%; Score 39; DB 11; Length 1734;
 Best Local Similarity 75.0%; Pred. No. 3.4e+02; Indels 0; Gaps 0;
 Matches 8; Conservative

OY 4 LLCPPRPL 11
 DB 139 LCPDPDPV 146
 |||||

Search completed: June 28, 2001, 11:55:32
 Job time: 153 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 66.55 seconds
(without alignments)
14.575 Million cell updates/sec

Title: US-09-439-313-554
Perfect score: 82
Sequence: 1 YVPLLEVGVEEKF 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	553	19	AAW71869
2	82	100.0	553	19	AAW69385
3	82	100.0	553	21	AAW28527
4	82	100.0	553	21	AAW82002
5	43	52.4	345	21	AAW11269
6	43	52.4	345	21	AAW51470
7	43	52.4	351	21	AAW11268
8	43	52.4	351	21	AAW51469
9	43	52.4	376	21	AAW11267
10	43	52.4	376	21	AAW51468
11	41	50.0	3472	21	AAW90913

ALIGNMENTS

RESULT	1
AAW71869	
ID	AAW71869 standard; Protein: 553 AA.
XX	
AC	AAW71869;
DT	
DT	06-JAN-1999 (first entry)
XX	
DE	Amino acid encoded by prostate tumour clone L1-12.
XX	
KW	Prostate; cancer; tumour; vaccine; immunogen; clone.
XX	
OS	Homo sapiens.
XX	

12	40	48.8	63	18	AAV11325	S. pneumoniae prot
13	40	48.8	163	21	AAV70541	Protein encoded by
14	40	48.8	172	19	AAV86095	S. pneumoniae deri
15	40	48.8	437	16	AAW82839	Sugar beet citrate
16	40	48.8	695	20	AAV20069	B. burgdorferi ant
17	40	48.8	718	20	AAV20068	B. burgdorferi ant
18	39	47.6	101	21	AAV79062	Polysaccharide bin
19	39	47.6	643	21	AAV87121	Human secreted pro
20	39	47.6	674	21	AAV87209	Human secreted pro
21	38	46.3	58	21	AAW44374	Human secreted pro
22	38	46.3	119	21	AAW42741	Human ORFX ORF2505
23	38	46.3	298	22	AAW50540	B. lactofermentum
24	38	46.3	312	22	AAW80121	Corynebacterium gl
25	37	45.1	119	11	AAW09304	Sequence deduced f
26	37	45.1	398	13	AAW20504	Mutant Human alpha
27	37	45.1	398	13	AAW20505	Mutant Human alpha
28	37	45.1	398	13	AAW20506	Mutant Human alpha
29	37	45.1	398	13	AAW20507	Mutant Human alpha
30	37	45.1	401	16	AAW83104	Alpha-1-antichymot
31	37	45.1	401	16	AAW83105	Alpha-1-antichymot
32	37	45.1	401	16	AAW83106	Alpha-1-antichymot
33	37	45.1	402	14	AAW44435	Alpha-1-antichymot
34	37	45.1	402	14	AAW42737	Alpha-1-antichymot
35	37	45.1	402	14	AAW42738	Alpha-1-antichymot
36	37	45.1	402	14	AAW42739	Alpha-1-antichymot
37	37	45.1	402	14	AAW42740	Alpha-1-antichymot
38	37	45.1	402	16	AAW83101	Wild-type alpha-1-
39	37	45.1	402	16	AAW83102	Alpha-1-antichymot
40	37	45.1	402	16	AAW83103	Alpha-1-antichymot
41	37	45.1	402	16	AAW67259	Alpha-1-antichymot
42	37	45.1	476	13	AAW20503	Human alpha-1-anti
43	37	45.1	476	14	AAW42736	Alpha-1-antichymot
44	37	45.1	476	16	AAW82250	Mature human wild
45	37	45.1	1410	21	AAW31539	Arabidopsis thalia
46	37	45.1	1455	21	AAW31538	Arabidopsis thalia
47	37	45.1	1465	21	AAW31537	Arabidopsis thalia
48	37	45.1	1693	12	AAW14618	Protein encoded by
49	36	43.9	83	21	AAW41140	Zea mays protein f
50	36	43.9	90	21	AAW41139	Zea mays protein f
51	36	43.9	209	22	AAW60348	Bovine beta-casein
52	36	43.9	307	20	AAW35623	Chlamydia pneumoni
53	36	43.9	353	21	AAW32571	Arabidopsis thalia
54	36	43.9	353	21	AAW49006	Arabidopsis thalia
55	36	43.9	354	21	AAW09010	Arabidopsis thalia
56	36	43.9	371	21	AAW32570	Arabidopsis thalia
57	36	43.9	371	21	AAW49005	Arabidopsis thalia
58	36	43.9	372	21	AAW09009	Arabidopsis thalia
59	36	43.9	458	21	AAW32569	Arabidopsis thalia
60	36	43.9	458	21	AAW49004	Arabidopsis thalia
61	36	43.9	459	21	AAW09008	Arabidopsis thalia
62	36	43.9	516	20	AAW55928	Human STLK3 protei
63	36	43.9	520	20	AAW01496	Human pan-s/tk-1A
64	36	43.9	520	20	AAW01497	Human pan-s/tk-1B
65	36	43.9	625	18	AAW09929	SEQ ID 2 from WO97

PT W09837093-A2.
 XX
 PD 27-AUG-1998.
 XX
 PF 25-FEB-1998; 98WO-US03492.
 XX
 PR 09-FEB-1999; 98US-0020956.
 XX
 PR 25-FEB-1997; 97US-0806093.
 XX
 PR 01-AUG-1997; 97US-0904804.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Dillon DC, Xu J;
 XX
 DR WPI: 1998-609886/51.
 XX
 DR N-PSDB; AAV61201.
 XX
 PT Polypeptides comprising immunogenic portions of prostate proteins
 XX used in a vaccine for the treatment of prostate cancer
 XX
 PS Example 1; Page 82-84; 130pp; English.
 CC The present sequence is an immunogenic portion of a prostate tumour
 CC protein. The immunogen, or the DNA encoding it, can be used as a
 CC vaccine for the treatment of prostate cancer. The immunogen was
 CC isolated from a prostate tumour cDNA library obtained by subtracting
 CC a prostate tumour cDNA expression library with a normal tissue cDNA
 CC library.
 XX
 XX Sequence 553 AA;
 SQ
 Query Match 100.0%; Score 82; DB 19; Length 553;
 Best Local Similarity 100.0%; Pred. NO. 3.7e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 YVPPLLLEVGVEKFM 16
 DB 38 YVPPLLLEVGVEKFM 53
 RESULT 2
 ID AAW69385 standard; Protein; 553 AA.
 AC AAW69385;
 XX
 DT 08-DEC-1998 (first entry)
 XX
 DE Prostate tumour specific gene clone L1-12 protein.
 XX
 KW Prostate tumour specific gene; human; prostate cancer; detection;
 KW therapy.
 XX
 OS Homo sapiens.
 XX
 PN W09837418-A2.
 XX
 PD 27-AUG-1998.
 XX
 PF 25-FEB-1998; 98WO-US03690.
 XX
 PR 09-FEB-1999; 98US-0904809.
 XX
 PR 25-FEB-1997; 97US-0806296.
 XX
 PR 01-AUG-1997; 97US-0904809.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Dillon DC, Xu J;
 XX
 DR WPI: 1998-480805/41.
 XX
 DR N-PSDB; AAV58586.
 XX

PT Novel human prostate specific tumour protein and fragments - useful
 XX for detecting and treating prostate cancers
 XX
 PS Example 1; Page 87-89; 141pp; English.
 XX
 CC This sequence is encoded by a human prostate tumour specific gene, and
 CC the present sequence is an immunogenic portion of a prostate tumour
 CC protein. The immunogen, or the DNA encoding it, can be used as a
 CC vaccine for the treatment of prostate cancer. The immunogen was
 CC isolated from a prostate tumour cDNA library obtained by subtracting
 CC a prostate tumour cDNA expression library with a normal tissue cDNA
 CC library.
 XX
 XX Sequence 553 AA;
 SQ
 Query Match 100.0%; Score 82; DB 19; Length 553;
 Best Local Similarity 100.0%; Pred. NO. 3.7e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 YVPPLLLEVGVEKFM 16
 DB 38 YVPPLLLEVGVEKFM 53
 RESULT 3
 ID AAB28527 standard; Protein; 553 AA.
 AC AAB28527;
 XX
 DT 07-FEB-2001 (first entry)
 XX
 DE Protein encoded by human breast tumour cDNA clone P501S.
 XX
 KW Human; breast tumour antigen; cytostatic; immunotherapy;
 KW breast cancer; vaccine.
 XX
 OS Homo sapiens.
 XX
 PN W0200061756-A2.
 XX
 PD 19-OCT-2000.
 XX
 PF 10-APR-2000; 2000WO-US09688.
 XX
 PR 09-APR-1999; 99US-0288950.
 XX
 PR 02-JUL-1999; 99US-0346327.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Reed SG, Xu J, Dillon DC;
 XX
 DR WPI: 2000-638568/61.
 XX
 DR N-PSDB; AAC79473.
 XX
 PT A novel isolated polypeptide comprising an immunogenic portion of a
 XX breast cancer protein useful in the detection and treatment of breast
 XX cancer -
 XX
 PS Claim 2; Page 92-93; 95pp; English.
 CC The present sequence is encoded by a cDNA sequence which was isolated
 CC from a breast tumour cDNA library. It is provided in a specification
 CC relating to compounds for immunotherapy and diagnosis of breast cancer.
 CC Breast tumour antigens and the polypeptides encoded by the cDNA
 CC sequence of a breast tumour cDNA library are provided in the
 CC treatment of breast cancer. Proliferated T cells and incubated antigen
 CC presenting cells are also required. The polypeptides and polynucleotides
 CC may also be used to produce a vaccine.
 XX

PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 23-JUN-1999; 99US-0139859.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141352.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 14-JUL-1999; 99US-0143242.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145111.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145951.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147292.
 PR 05-AUG-1999; 99US-0147292.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 09-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 11-AUG-1999; 99US-0148119.
 PR 12-AUG-1999; 99US-0148119.
 PR 13-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149369.
 PR 17-AUG-1999; 99US-0149125.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0150048.
 PR 25-AUG-1999; 99US-0150526.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151303.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154792.
 PR 22-SEP-1999; 99US-0155113.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157805.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159538.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160983.
 PR 22-OCT-1999; 99US-0161044.
 PR 22-OCT-1999; 99US-0161405.
 PR 23-OCT-1999; 99US-0161406.
 PR 23-OCT-1999; 99US-0161329.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 28-OCT-1999; 99US-0161993.
 PR 28-OCT-1999; 99US-0162142.

Query Match 52.4%; Score 43; DB 21; Length 345;
 Rest Local Similarity 61.5%; Pred No. 13;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 QY 3 PPLLEVGVEEKF 15
 Db 118 PpIlIqgqeef 130
 RESULT 6
 AAG51470

ID AAG51470 standard; Protein; 345 AA.
XX AAG51470;
AC AAG51470;
XX AAG51470;
XX AAG51470;
DT 18-OCT-2000 (first entry)
XX AAG51470;
DE Arabidopsis thaliana protein fragment SEQ ID NO: 65328.
XX Arabidopsis thaliana protein fragment SEQ ID NO: 65328.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX Arabidopsis thaliana.
PN EP1033405-A2.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PF 25-FEB-1999; 99US-0121825.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132486.
XX 11-MAY-1999; 99US-0132486.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0134941.
XX 21-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139452.
XX 17-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0140353.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145224.
PR 28-JUL-1999; 99US-0145276.
PR 28-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.


```
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0143390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153358.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.4%; Score 43; DB 21; Length 351;
Best Local Similarity 61.5%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLILEVGVBEKF 15
   |||::|||
Db 124 ppilqngqeenf 136

RESULT 8
AAG51469
ID AAG51469 standard; Protein; 351 AA.
XX
AC AAG51469;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 65327.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
```

[illegible]

PR	20-SEP-1999;	99US-0154779.	PR	05-MAR-1999;	99US-0123180.
PR	22-SEP-1999;	99US-0155139.	PR	09-MAR-1999;	99US-0123548.
PR	23-SEP-1999;	99US-0155486.	PR	23-MAR-1999;	99US-0125788.
PR	24-SEP-1999;	99US-0155659.	PR	25-MAR-1999;	99US-0126264.
PR	28-SEP-1999;	99US-0156458.	PR	29-MAR-1999;	99US-0126785.
PR	29-SEP-1999;	99US-0156596.	PR	01-APR-1999;	99US-0127462.
PR	04-OCT-1999;	99US-0157117.	PR	06-APR-1999;	99US-0128234.
PR	05-OCT-1999;	99US-0157753.	PR	08-APR-1999;	99US-0128714.
PR	06-OCT-1999;	99US-0157865.	PR	16-APR-1999;	99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159330.	PR	04-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999;	99US-0132486.
PR	14-OCT-1999;	99US-0159638.	PR	06-MAY-1999;	99US-0132487.
PR	18-OCT-1999;	99US-0159584.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160741.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160767.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134370.
PR	21-OCT-1999;	99US-0160815.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160980.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160981.	PR	20-MAY-1999;	99US-0135124.
PR	22-OCT-1999;	99US-0160989.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161404.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161405.	PR	25-MAY-1999;	99US-0136021.
PR	25-OCT-1999;	99US-0161406.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161359.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161360.	PR	01-JUN-1999;	99US-0137222.
PR	26-OCT-1999;	99US-0161361.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-016161920.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161992.	PR	07-JUN-1999;	99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999;	99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138540.
PR	29-OCT-1999;	99US-0162143.	PR	10-JUN-1999;	99US-0138847.
PR	29-OCT-1999;	99US-0162144.	PR	14-JUN-1999;	99US-0139119.
PR	29-OCT-1999;	99US-0162145.	PR	16-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162146.	PR	16-JUN-1999;	99US-0139453.
PR	29-OCT-1999;	99US-0162147.	PR	17-JUN-1999;	99US-0139492.
PR	29-OCT-1999;	99US-0162148.	PR	18-JUN-1999;	99US-0139454.
PR	29-OCT-1999;	99US-0162149.	PR	18-JUN-1999;	99US-0139455.
PR	29-OCT-1999;	99US-0162150.	PR	18-JUN-1999;	99US-0139456.
PR	29-OCT-1999;	99US-0162151.	PR	18-JUN-1999;	99US-0139457.
PR	29-OCT-1999;	99US-0162152.	PR	18-JUN-1999;	99US-0139458.
PR	29-OCT-1999;	99US-0162153.	PR	18-JUN-1999;	99US-0139459.
PR	29-OCT-1999;	99US-0162154.	PR	18-JUN-1999;	99US-0139460.
PR	29-OCT-1999;	99US-0162155.	PR	18-JUN-1999;	99US-0139461.
PR	29-OCT-1999;	99US-0162156.	PR	18-JUN-1999;	99US-0139462.
PR	29-OCT-1999;	99US-0162157.	PR	18-JUN-1999;	99US-0139463.
PR	29-OCT-1999;	99US-0162158.	PR	18-JUN-1999;	99US-0139750.
PR	29-OCT-1999;	99US-0162159.	PR	18-JUN-1999;	99US-0139763.
PR	29-OCT-1999;	99US-0162160.	PR	21-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162161.	PR	22-JUN-1999;	99US-0139899.
PR	29-OCT-1999;	99US-0162162.	PR	23-JUN-1999;	99US-0140353.
PR	29-OCT-1999;	99US-0162163.	PR	23-JUN-1999;	99US-0140354.
PR	29-OCT-1999;	99US-0162164.	PR	24-JUN-1999;	99US-0140695.
PR	29-OCT-1999;	99US-0162165.	PR	28-JUN-1999;	99US-0140823.
PR	29-OCT-1999;	99US-0162166.	PR	29-JUN-1999;	99US-0140991.
PR	29-OCT-1999;	99US-0162167.	PR	30-JUN-1999;	99US-0141287.
PR	29-OCT-1999;	99US-0162168.	PR	01-JUL-1999;	99US-0141842.
PR	29-OCT-1999;	99US-0162169.	PR	01-JUL-1999;	99US-0142154.
PR	29-OCT-1999;	99US-0162170.	PR	02-JUL-1999;	99US-0142055.
PR	29-OCT-1999;	99US-0162171.	PR	06-JUL-1999;	99US-0142390.
PR	29-OCT-1999;	99US-0162172.	PR	08-JUL-1999;	99US-0142803.
PR	29-OCT-1999;	99US-0162173.	PR	09-JUL-1999;	99US-0142920.
PR	29-OCT-1999;	99US-0162174.	PR	12-JUL-1999;	99US-0142977.
PR	29-OCT-1999;	99US-0162175.	PR	13-JUL-1999;	99US-0143342.

Query Match 52.4%; Score 43; DB 21; Length 351;
Best Local Similarity 61.5%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLLEVGVEEF 15
|||:|:|
Db 124 ppllpggqeenf 136

RESULT 9
AAG11267
ID AAG11267 standard; Protein; 376 AA.
XX AC AAG11267;
XX AC AAG11267;
DT 17-OCT-2000 (first entry)
XX XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9920.
XX XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX XX
OS Arabidopsis thaliana.
XX XX
PN EP1033405-A2.
XX XX
PD 06-SEP-2000.
XX XX
PF 25-FEB-2000; 2000EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144634.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.

```

PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159784.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.4%; Score 43; DB 21; Length 376;
Best Local Similarity 61.5%; Pred No 15;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 PPLLEVGVEEF 15
DB 149 ppllgnqgeenf 161

RESULT 11
ID AAY90913 standard; Protein; 3472 AA.
XX
AC AAY90913;
DT 30-AUG-2000 (first entry)
XX
DE Cenarchaeum symbiosum open reading frame protein sequence SEQ ID NO:4.
XX
KW Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
KW Characterisation; archae; therapeutic; industrial; laboratory.
XX
OS Cenarchaeum symbiosum.
XX
PN WO200018909-A2.
XX
PD 06-APR-2000.
XX
XX 29-SEP-1999; 99WO-US22752.
XX
XX 29-SEP-1998; 98US-0102294.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Swanson RV, Feldman RA, Schleper C;
XX WPI: 2000-293148/25.
DR N-PSDB; AAA55188.
XX
XX New nucleic acids and proteins isolated from the non-thermophilic
XX crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
XX physiology of these archae and in therapeutic, industrial or laboratory
XX techniques.
PS Claim 26; Page 102-110; 210pp; English.

```

```

CC AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids
CC and proteins isolated from the non-thermophilic crenarchaeote
CC Cenarchaeum symbiosum. The nucleic acids and proteins identified in
CC the present invention are useful in characterizing the physiology of
CC these archae and can be used in therapeutic, industrial or laboratory
CC techniques. AAA55227 to AAA55260 represent promoter sequences from
CC Cenarchaeum symbiosum. AAA55261 to AAA55286 represent PCR primers and
CC probes used in examples. From the present invention.
XX
SQ Sequence 3472 AA;

Query Match 50.0%; Score 41; DB 21; Length 3472;
Best Local Similarity 66.7%; Pred No 3;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 VPPLLEVGVEE 13
DB 1169 lppflfigvgae 1180

RESULT 12
ID AAY11325 standard; Protein; 63 AA.
XX
AC AAY11325;
XX
XX 20-MAY-1999 (first entry)
XX
DE S. pneumoniae protein SEQ ID NO:435.
XX
KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;
KW streptococcal infection; pneumococcal.
XX
XX Streptococcus pneumoniae.
XX
PN WO9737026-A1.
XX
PD 09-OCT-1997.
XX
XX 01-APR-1997; 97WO-US05306.
XX
PR 22-AUG-1996; 96US-0025788.
PR 02-APR-1996; 96US-0014690.
XX
XX (SMIK ) SMITHKLINE BEECHAM CORP.
PA (SMIK ) SMITHKLINE BEECHAM PLC.
XX
XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
PI Stodola RK;
XX
XX WPI: 1997-503111/46.
DR N-PSDB; AAX30907.
XX
XX Nucleic acids encoding pneumococcal polypeptide(s) - useful in
XX vaccines, drug screening, etc
XX
XX Claim 6; Page 320-321; 354pp; English.
XX
XX AAX30724 to AAX30946 represent genomic DNA sequences isolated from
XX Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
XX include the novel pneumococcal protein-coding regions. The
XX isolated from Streptococcus pneumoniae can be used in vaccines against
XX streptococcal infections and in assays for identifying compounds that
XX inhibit or activate the activity of the proteins. The antagonists can
XX be used to treat an individual having need to inhibit a bacterial
XX protein. Vectors expressing the proteins can be used to induce a
XX protective immune response in mammals.
XX
XX Sequence 63 AA;

Query Match 48.8%; Score 40; DB 18; Length 63;

```

Best Local Similarity 42.9%; Pred. No. 7.1;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEK 14
Db 36 Yippvvlndvdpnk 49
|:|:|:| | : |

RESULT 13
AAAY70541
ID AAY70541 standard; Protein; 163 AA.
XX
AC AAY70541;
XX
DT 04-JUL-2000 (first entry)
XX
DE Protein encoded by N. tabacum Enh2 gene for enhancing INFL-induced HR.
XX
KW Enh2; enhancer of INFL-induced hypersensitive response; tobacco;
KW resistance gene; R gene; non-host disease resistance; plant pathogen;
KW hypersensitive response; HR; Phytophthora infestans elicitor; INFL;
KW transgenic plant; viral pathogen; fungal pathogen; bacterial pathogen;
KW nematodal pathogen; Phytophthora infestans.
XX
OS Nicotiana tabacum.
XX
PH Key Location/Qualifiers
FT Misc-difference 159
FT /note= "Encoded by tag"
XX
PN WO200012736-A2.
XX
PD 09-MAR-2000.
XX
PF 31-AUG-1999; 99WO-US19899.
XX
PR 31-AUG-1998; 98US-0098402.
XX
PA (MONS) MONSANTO CO.
XX
PI Rommens CMT, Swords KMM, Yan H, Zhang B;
XX
DR WPI; 2000-256651/22.
DR N-PSDB; AAZ51874.
XX
PT Identification of non-host plant disease resistance genes comprises
PT expressing resistance and non-host inducible genes in susceptible
PT plants -
XX
PS Claim 15; Page 69; 94pp; English.
XX
CC The patent discloses a method for identifying genes that enhance levels
CC of disease resistance if expressed in susceptible plants. The method is
CC useful for isolating disease resistance genes (R gene) in plants. These
CC genes confer non-host disease resistance to plants by responding to
CC avirulence genes in plant pathogens. The R-genes identified trigger a
CC hypersensitive response (HR) in tobacco that is dependent on the presence
CC of the Phytophthora infestans elicitor INFL. The genes are useful for
CC generating pathogen-resistant transgenic plants. The transgenic plants
CC are preferably Acacia, apple, banana, barley, bean, broccoli, cabbage,
CC canola, carrot, citrus, coffee, corn, cotton, cucumber, Douglas fir,
CC Eucalyptus, garlic, grape, loblolly pine, melon, oat, oil palm, onion, an
CC ornamental plant, pea, peanut, pepper, Poplar tree, potato, Radiata pine,
CC rice, rye, sorghum, Southern pine, soybean, strawberry, sugarbeet,
CC sugarcane, sunflower, Sweetgum, tea, tomato, turf, a vine and wheat. They
CC can be used to control viral, fungal, bacterial or nematodal pathogens,
CC e.g. Phytophthora, Erysiphe and Puccinia. The present sequence is encoded
CC by partial sequence of Enh2 (enhancer of INFL-induced HR) tobacco R gene
CC that enhances the INFL-dependent HR in Nicotiana benthamiana.
XX
SQ Sequence 163 AA;

Query Match 48.8%; Score 40; DB 21; Length 163;
Best Local Similarity 58.3%; Pred. No. 20;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 PPLLEVGVEEK 14
Db 149 ppeildvgfezq 160
|:|:|:| | : |

RESULT 14
AAAY86095
ID AAY86095 standard; Protein; 172 AA.
XX
AC AAY86095;
XX
DT 10-APR-2000 (first entry)
XX
DE S. pneumoniae derived protein #304.
XX
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection.
XX
OS Streptococcus pneumoniae.
XX
PN WO9806734-A1.
XX
PD 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US14436.
XX
PR 16-AUG-1996; 96US-0024022.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;
XX
DR WPI; 1998-159452/14.
DR N-PSDB; AAZ96414.
XX
PT Streptococcus pneumoniae proteins and related DNA - useful for
PT screening compounds for antibacterial activity
XX
PS Claim 5; Page 571; 640pp; English.
XX
CC This invention describes novel isolated Streptococcus pneumoniae
CC polynucleotides (see AAZ96173-296494) and their encoded proteins (see
CC AAY8592-Y86182). The DNA, vectors and host cells described in the
CC method of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The
CC polypeptides (or DNA encoding them, via gene therapy) are also useful
CC for inducing an immunological response in a mammal. The antagonists are
CC useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease.
XX
SQ Sequence 172 AA;

Query Match 48.8%; Score 40; DB 19; Length 172;
Best Local Similarity 42.9%; Pred. No. 21;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEK 14
Db 145 Yippvvlndvdpnk 158
|:|:|:| | : |

RESULT 15

AA82839
 ID AAR82839 standard; Protein; 437 AA.
 AC AAR82839;
 XX
 XX 25-JAN-1996 (first entry)
 XX
 XX Sugar beet citrate synthase.
 XX
 XX Citrate synthase; flower formation.
 XX
 XX Beta vulgaris strain Zuchtlinie 5S 0026.
 XX
 XX WO9524487-A.
 XX
 XX 14-SEP-1995.
 XX
 XX 07-MAR-1995; 95WO-EP00859.
 XX
 XX 19-OCT-1994; 94DE-4438821.
 XX
 XX 09-MAR-1994; 94DE-4408629.
 XX
 XX 22-SEP-1994; 94DE-4435566.
 XX
 XX (AGRE) HOECHST-SCHERING AGREVO GMBH.
 XX
 XX La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 XX
 XX Mueller-roeber B;
 XX
 XX WPI; 1995-328278/42.
 XX
 XX N-PSDB; AAT04200.
 XX
 XX DNA encoding plant citrate synthase - used to regulate flower formation,
 XX
 XX to improve storage of tubers, etc. and to reduce sprouting
 XX
 XX Disclosure; Page 57-60; 87pp; English.
 XX
 XX To identify a cDNA from sugar beet which codes for citrate
 XX
 XX synthase, a cDNA bank of leaf tissue from sugar beet was prepd.
 XX
 XX Plaques of this cDNA bank were screened using radioactive DNA
 XX
 XX probes which comprise a mixture of Solanum tuberosum citrate
 XX
 XX synthase cDNA (AA04199), and Nicotian tabacum citrate synthase cDNA
 XX
 XX (AA04200). One of the clones was sequenced. The nt. sequence is
 XX
 XX given in AAT04200.
 XX
 XX Sequence 437 AA;

Query Match 48.8%; Score 40; DB 16; Length 437;
 Query Local Similarity 77.9%; Protein 57;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPPLLEVG 10
 DB 356 vppillelg 364

RESULT 16
 AAY20069
 ID AAY20069 standard; Protein; 695 AA.
 AC AAY20069;
 XX
 XX 19-JUL-1999 (first entry)
 XX
 XX B. burgdorferi antigenic protein, t679.aa.
 XX
 XX Antigenic protein; vaccine; Lyme disease; infection; detection.
 XX
 XX Borrelia burgdorferi.
 XX
 XX WO9859071-A1.
 XX
 XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
 XX
 XX WPI; 1999-189980/16.
 XX
 XX 30-DEC-1998.

XX 18-JUN-1998; 98WO-US12718.
 XX
 XX 03-SEP-1997; 97US-0057483.
 XX
 XX 20-JUN-1997; 97US-0050359.
 XX
 XX 22-JUL-1997; 97US-0053344.
 XX
 XX 22-JUL-1997; 97US-0053377.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX (MEDI-) MEDIMUNE INC.
 XX
 XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
 XX
 XX WPI; 1999-189980/16.
 XX
 XX N-PSDB; AAX01766.
 XX
 XX New isolated Borrelia burgdorferi nucleic acids - used to develop
 XX
 XX products for the diagnosis, prevention and treatment of diseases
 XX
 XX caused by Borrelia, particularly Lyme disease
 XX
 XX Claim 12; Page 183; 275pp; English.
 XX
 XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
 XX
 XX invention, which is suitable for use in a vaccine. The Bb polypeptides
 XX
 XX can be used in vaccines for eliciting protective antibodies to members of
 XX
 XX the Borrelia genus, particularly for the use against Lyme disease in
 XX
 XX humans and animals. They can be used for preventing or attenuating an
 XX
 XX infection caused by a member of the Borrelia genus. The products can also
 XX
 XX be used for detection of members of the Borrelia genus.
 XX
 XX Sequence 695 AA;
 XX
 XX Query Match 48.8%; Score 40; DB 20; Length 695;
 XX
 XX Query Local Similarity 50.0%; Protein 94;
 XX
 XX Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPLLEVGVERKFM 16
 DB 51 Yvaylfkkigfeekiv 66

RESULT 17
 AAY20068
 ID AAY20068 standard; Protein; 718 AA.
 AC AAY20068;
 XX
 XX 19-JUL-1999 (first entry)
 XX
 XX B. burgdorferi antigenic protein, f679.aa.
 XX
 XX Antigenic protein; vaccine; Lyme disease; infection; detection.
 XX
 XX Borrelia burgdorferi.
 XX
 XX WO9859071-A1.
 XX
 XX 30-DEC-1998.
 XX
 XX 18-JUN-1998; 98WO-US12718.
 XX
 XX 03-SEP-1997; 97US-0057483.
 XX
 XX 20-JUN-1997; 97US-0050359.
 XX
 XX 22-JUL-1997; 97US-0053344.
 XX
 XX 22-JUL-1997; 97US-0053377.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX (MEDI-) MEDIMUNE INC.
 XX
 XX Choi GH, Erwin AL, Hanson MS, Lathigra R;
 XX
 XX WPI; 1999-189980/16.

DR N-PSDB; AAX61765.
XX
XX New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX
XX Claim 12; Page 183; 275pp; English.
XX
XX This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.
XX
XX Sequence 718 AA;
SQ

Query Match 48.8%; Score 40; DB 20; Length 718;
Best Local Similarity 50.0%; Pred. No. 98;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 YVPPLLLEVGVEEKFM 16
Db 74 yvaylfkkigfeekfv 89
II I :|||II|

RESULT 18
AAY79062
ID AAY79062 standard; peptide; 101 AA.
XX
XX AAY79062;
AC
DT 12-JUN-2000 (first entry)
DE
DE Polysaccharide binding domain of human PPp1R6.
XX
XX Protein phosphatase 1; Pp1; glycogen targetting subunit; blood glucose;
KW phosphorylase alpha; glycogen synthesis; hyperglycaemic disorder;
KW type I diabetes; type II diabetes; polysaccharide binding domain; PPp1R6.
XX
XX Homo sapiens.
OS
PN WO200012549-A1.
XX
XX 09-MAR-2000.
PD
XX 19-AUG-1999; 99WO-GB02761.
PF
XX 27-AUG-1998; 98GB-0018650.
PR
XX (MEDI-) MEDICAL RES COUNCIL.
PA
XX Cohen PTW, Armstrong CG, Doherty MJ;
PI
XX WPI; 2000-256587/22.
DR
XX Lowering blood sugar levels in the treatment of diabetes, using a
PT compound that blocks interaction between phosphorylase alpha and
PT protein phosphatase 1 glycogen-targetting subunit
XX
XX Disclosure; Fig 6; 51pp; English.
XX

This sequence represents the amino acid sequence of the human
CC polysaccharide targetting domain of the human protein phosphatase 1
CC glycogen targetting subunit PPp1R6. The invention relates to the
CC medicinal use of a compound capable of blocking the interaction of
CC phosphorylase alpha with GL, where the compound comprises the C-terminal
CC peptide of rat GL (see AAY79064) or a fragment of it. When phosphorylase
CC alpha binds to GL it potentially inhibits its glycogen synthase phosphatase
CC activity and inhibits glycogen synthesis. This contributes to high blood
CC glucose levels. The invention also relates to a method for identifying

the compound. The compound can be used to reduce the blood glucose level
of a mammal, particularly a human, in hyperglycaemic disorders such as
type I or type II diabetes.
Sequence 101 AA;
Query Match 47.6%; Score 39; DB 21; Length 101;
Best Local Similarity 57.1%; Pred. No. 18;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 2 VPPLLLEVGVEEKFM 15
Db 66 vppflllelgsrvhf 79
||| ||||| I

RESULT 19
AAY87121
ID AAY87121 standard; Protein; 643 AA.
XX
XX AAY87121;
AC
XX 09-MAY-2000 (first entry)
DT
DE Human secreted protein sequence SEQ ID NO:160.
DE
XX Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiinflammatory; nootropic; neuroprotective; antiallergic; cancer;
KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
KW foetal deficiency; blood disorder; immune system disorder; arthritis;
KW autoimmune disease; hepatic disease; renal disease; inflammation;
KW Alzheimer's disease; behavioural disorder; schizophrenia; osteoporosis;
KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
KW metabolic disorder; food additive; preservative.
XX
XX Homo sapiens.
OS
XX WO200004140-A1.
PN
XX 27-JAN-2000.
PD
XX 14-JUL-1999; 99WO-US15849.
PF
XX 15-JUL-1998; 98US-0092921.
PR
XX 15-JUL-1998; 98US-0092922.
PR
XX 15-JUL-1998; 98US-0092956.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Ebner R, Olsen HS, Brewer LA, Florence KA, Young PE;
PI Mucenski M, Endress GA, Soppet DR;
PI
XX WPI; 2000-161128/14.
DR
XX N-PSDB; AAY87121.
DR
XX New isolated human genes, useful for diagnosis and treatment of, e.g.
PT cancers, neurological or blood disorders
PT
XX Claim 11; Page 415-417; 494pp; English.
XX

The polynucleotide sequences given in AAY87121 to AAY87123 encode the
CC human secreted proteins given in AAY87064 to AAY87223. Human secreted
CC protein can have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: cytostatic;
CC immunosuppressive; antiinflammatory; nootropic; neuroprotective; and
CC antiallergic. The polynucleotides and their corresponding secreted
CC polypeptides are useful for preventing, treating or ameliorating medical
CC conditions, e.g. by protein or gene therapy. Also pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new

CC polynucleotides. Human secreted protein s and their polynucleotides can
 CC be in a variety of diseases, including cancer,
 CC tumours, neurodegenerative disorders, developmental abnormalities and
 CC foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal diseases, inflammation,
 CC allergies, Alzheimer's disease, behavioural disorders, schizophrenia,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC disorders, respiratory disorders and metabolic disorders. The
 CC proteins or polynucleotides can also be used as food additives or
 CC preservatives. The proteins are also useful for identifying their
 CC binding partners. AA298008 to AA298016 and AA87063 are sequence used in
 CC the exemplification of the present invention.

SQ Sequence 643 AA;

Query Match 47.6%; Score 39; DB 21; Length 643;
 Best Local Similarity 60.0%; Pred. No. 1.3e+02;
 Matches 9; Conservative 2; Mismatches 2; Indels 1;

OY 2 VPPLLEVGVEKFM 16

Db 270 vppllcmmg--ekfl 282

RESULT 20

AA298008 standard; Protein: 674 AA.

XX AA87209;

XX AA87209;

XX 09-MAY-2000 (first entry)

XX Human secreted protein sequence SEQ ID NO:248.

KW Human; secreted protein; diagnosis: cytostatic; immunosuppressive;
 KW antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;
 KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
 KW foetal deficiency; blood disorder; immune system disorder; arthritis;
 KW autoimmune disease; hepatic disease; renal disease; inflammation; osteoporosis;
 KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
 KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
 KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
 KW metabolic disorder; food additive; preservative.

XX Homo sapiens.

XX WO200004140-A1.

XX 27-JAN-2000.

XX 14-JUL-1999; 99WO-US15849.

XX 15-JUL-1998; 98US-0092921.

XX 15-JUL-1998; 98US-0092922.

XX 15-JUL-1998; 98US-0092956.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis G, Duan RD, Rosen CA, Moore PA, Shi Y;

XX Kufundwa, Olsan, Brewster LA, Florence RA, Young PE;

XX Mucenski M, Endress GA, Soppet DR;

XX WPI; 2000-161128/14.

XX N-PSDB; AA298074.

XX New isolated human genes, useful for diagnosis and treatment of, e.g.

XX Cancers, neurological or blood disorders

XX Disclosure; Page 479-482; 494pp; English.

PS

XX The polynucleotide sequences given in AA298017 to AA298108 encode the
 CC human secreted proteins given in AA87064 to AA87223. Human secreted
 CC protein can have activities based on the tissues and cells the genes are
 CC expressed in. Examples of activities include: cytostatic;
 CC immunosuppressive; antiinflammatory; nontropic; neuroprotective; and
 CC antiallergic. The polynucleotides and their corresponding secreted
 CC polypeptides are useful for preventing or treating or diagnosing medical
 CC conditions. The polynucleotides and their corresponding secreted
 CC polypeptides can be used to determine the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Human secreted protein s and their polynucleotides can
 CC be used for developing products for the diagnosis or treatment of cancer,
 CC tumours, neurodegenerative disorders, developmental abnormalities and
 CC foetal deficiencies, hepatic and renal disease, inflammation, system,
 CC osteoporosis, arthritis, infections, AIDS, spinal cord injuries,
 CC transplant rejection, diabetes, asthma, sepsis, acne, psoriasis,
 CC cardiovascular disorders, reproductive disorders, gastrointestinal
 CC disorders, respiratory disorders and metabolic disorders. The
 CC proteins or polynucleotides can also be used as food additives or
 CC preservatives. The proteins are also useful for identifying their
 CC binding partners. AA298008 to AA298016 and AA87063 are sequence used in
 CC the exemplification of the present invention.

XX Sequence 674 AA;

Query Match 47.6%; Score 39; DB 21; Length 674;

Best Local Similarity 60.0%; Pred. No. 1.4e+02;

Matches 9; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

OY 2 VPPLLEVGVEKFM 16

Db 301 vppllcmmg--ekfl 313

RESULT 21

ID AAB44374 standard; Protein: 58 AA.

XX AAB44374;

XX AAB44374;

XX 14-FEB-2001 (first entry)

XX Human secreted protein encoded by gene 41 clone HSIDA39.

XX Cytostatic; immunosuppressive; nontropic; neuroprotective; antiviral;
 KW antiinflammatory; nontropic; neuroprotective; antiallergic; cancer;
 KW tumour; neurodegenerative disorder; developmental abnormality; allergy;
 KW foetal deficiency; blood disorder; immune system disorder; arthritis;
 KW autoimmune disease; hepatic disease; renal disease; inflammation; osteoporosis;
 KW infection; AIDS; spinal cord injury; transplant rejection; diabetes;
 KW asthma; sepsis; acne; psoriasis; cardiovascular disorder;
 KW reproductive disorder; gastrointestinal disorder; respiratory disorder;
 KW metabolic disorder; food additive; preservative.

XX Homo sapiens.

XX WO2000058358-A1.

XX 05-OCT-2000.

XX 23-MAR-2000; 2000WO-US07725.

XX 26-MAR-1999; 98US-0126502.

XX 14-JAN-2000; 2000US-0176063.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX WPI; 2000-594540/56.

XX N-PSDB; AAC79037.

XX Forty nine nucleic acid molecules encoding human secreted proteins,

PT useful in the prevention, treatment and diagnosis of cancer, immune disorders, cardiovascular disorders and neurological diseases -

XX Claim 11; Page 353; 367pp; English.

XX Sequences AAB44335-B44382 represent the amino acid sequences of 49 human secreted proteins encoded by the genes AAC69084-C69119. The genes and proteins are useful for preventing, ameliorating or treating medical conditions, e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

XX Sequence 58 AA;

Query Match 46.3%; Score 38; DB 21; Length 58;
Best Local Similarity 81.8%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLELVGVEEKF 15
||||| |||||
Db 12 lllepvaeekf 22

RESULT 22
AAB42741
ID AAB42741 standard; Protein; 119 AA.

AC AAB42741;

XX 08-FEB-2001 (first entry)

Human ORFX ORF2505 polypeptide sequence SEQ ID NO:5010.

DE Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnerable; antiparotrophic; antiparkinsonian; nontropic; neuroprotective; anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antiviral; antibacterial; antifungal; antirheumatic; antithyroid; antianaemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive.

OS Homo sapiens.

XX WO200058473-A2.

PN 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

XX (CURA-) CUPAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI: 2000-602362/57.

DR N-PSDB; AAC76950.

XX Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -

PT Claim 11; Page 4188; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human ORFX open reading frames 1 to 3161. The ORFX sequences have activities such as: cytostatic; hepatotropic; vulnerable; antiparotrophic; antiparkinsonian; nontropic; neuroprotective; osteopathic; anticonvulsant; antiarthritic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive; dermatological; immunosuppressive; antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The sequences can be used for determining the presence of or predisposition to, or preventing or treating pathological conditions associated with an ORFX-associated disorder. The nucleic acids can be used to express ORFX proteins in gene therapy vectors. The proteins and nucleic acids may be used to treat cancers, proliferative disorders, neurodegenerative disorders, osteoarthritis, graft vs host disease, cardiovascular disease, diabetes mellitus, hypertension, hypothyroidism, cholesterol ester storage, systemic lupus erythematosus, severe combined immunodeficiency (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 119 AA;

Query Match 46.3%; Score 38; DB 21; Length 119;
Best Local Similarity 57.1%; Pred. No. 31;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 PLLLLLEGVVEEKFM 16
| ||| | ||| :
Db 71 pnlllpfgteekll 84

RESULT 23
AAB50540
ID AAB50540 standard; Protein; 298 AA.

XX AAB50540;

AC AAB50540;

XX 20-MAR-2001 (first entry)

DT B. lactofermentum pyrimidine biosynthetic system enzyme protein #1.

DE Brevibacterium lactofermentum; coryneform; microbe; dihydroorotase; pyrimidine biosynthetic system; aspartate carbamoyltransferase; orotidylate decarboxylase.

XX Brevibacterium lactofermentum.

OS JP2000287688-A.

PN 17-OCT-2000.

XX 08-APR-1999; 99JP-0100989.

PF 08-APR-1999; 99JP-0100989.

PR (AJIN) AJINOMOTO KK.

XX WPI: 2001-065574/08.

DR N-PSDB; AAC90812.
 XX A Pyrimidine biosynthetic system enzyme gene of a coryneform microbe
 PT Claim 1; Page 18-19; 25pp; Japanese.
 PS
 CC The present invention describes a pyrimidine biosynthetic system enzyme
 CC gene of a coryneform microbe e.g. Brevibacterium lactofermentum. Also
 CC defined 298 amino acid sequence (P1) (AAB80540) or a sequence in which
 CC at least one amino acid residue is replaced, deleted, inserted, added or
 CC reversed in P1, where the protein has aspartate carbamoyltransferase
 CC activity; (2) a DNA encoding a protein comprising a defined 397 amino
 CC acid sequence (P2) (AAB80541) or a sequence in which at least one amino
 CC acid residue is replaced, deleted, inserted, added or reversed in P2,
 CC where the protein has dihydroorotase activity; (3) a DNA encoding a
 CC protein comprising a defined 278 amino acid sequence (P3) (AAB80548) or a
 CC sequence in which at least one amino acid residue is replaced, deleted,
 CC inserted, added or reversed in P3, where the protein has orotidylate
 CC decarboxylase activity; (4) a coryneform microbe transformed by the DNA;
 CC gene (3) the proteins, peptides and their variants encoding by the DNA. The
 CC present sequence represents a protein encoded by a chromosome DNA
 CC fragment from Brevibacterium lactofermentum (AAC90812), which is a
 CC pyrimidine biosynthetic system enzyme.
 XX Sequence 298 AA:
 SQ

Query Match 46.3%; Score 38; DB 22; Length 298;
 Best Local Similarity 70.0%; Pred. NO. 84;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0:
 QY 3 PRLLEVGVE 12
 DB 178 Pptllpigr 187

RESULT 24
 AAB80121
 XD AAB80121 standard; Protein; 312 AA.
 XX
 AC AAB80121;
 XX
 DT 30-APR-2001 (first entry)
 XX
 XX Corynebacterium glutamicum MP protein sequence SEQ ID NO:976.
 DE
 XX Corynebacterium glutamicum; metabolic pathway protein: MP protein;
 XX fine chemical production; microorganism: organic acid; nucleoside;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
 XX
 XX Corynebacterium glutamicum.
 OS
 PN WO200100843-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00923.
 XX
 XX 25-JUN-1999; 99US-0141031.
 PR 01-JUL-1999; 99DE-1030476.
 PR 02-JUL-1999; 99US-0142101.
 PR 08-JUL-1999; 99DE-1031415.
 PR 08-JUL-1999; 99DE-1031418.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031421.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031435.

PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031563.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032123.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 27-AUG-1999; 99DE-1040744.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041384.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 XX
 XX (BADI) BASF AG.
 XX
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 XX WPI; 2001-137957/14.
 DR N-PSDB; AAF72240.
 XX
 XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
 XX pathway proteins useful for producing fine chemicals in
 PT microorganisms including organic acids, nonproteinogenic amino acids,
 PT and purine and pyrimidine bases -
 XX
 XX Claim 20; Page 1496-1497; 1737pp; English.
 PS
 XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
 CC pathway (MP). Proteins given in AAF79634 to AAB80211 the C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.
 XX
 XX Sequence 312 AA;
 SQ

Query Match 46.3%; Score 38; DB 22; Length 312;
Best Local Similarity 70.0%; Pred. No. 89;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 PPLLEVGVE 12
DB 192 pptllpvgve 201

RESULT 25

AAR09304
ID AAR09304 standard; protein; 119 AA.

XX AC AAR09304;

DT 27-FEB-1991 (first entry)

XX DE Sequence deduced from rev gene of HIV 1-NDK.

XX KW Human immunodeficiency virus; AIDS.

XX OS HIV 1-NDK.

XX FH Key Location/Qualifiers
FT Region 1..30
FT /label= rev ex2
FT Region 31..119
FT /label= rev ex3

XX PN W09013630-A.

XX PD 15-NOV-1990.

XX PF 02-MAY-1990; 90WO-FR00312.

XX PR 03-MAY-1989; 89FR-0005914.

XX PA (INRM) INSERM INST NAT SANTE.

XX PI Barre-Sinoussi F, Chermann JC, Devaux C, Rey F, Sire J;
PI Spire B;

XX DR WPI; 1990-361470/48.

XX DR N-PSDB; AAR06635.

XX PT New HIV-NDK retrovirus and protein component - used in vaccines
PT against immuno-deficiency disorders and in raising MAbs for
PT retro-virus detection in vivo.

XX PS Disclosure; Fig 2; 37pp; French.

XX CC The HIV NDK virus was isolated from peripheral blood lymphocytes of
CC an AIDS patient. A genomic library was prep'd. from DNA extracted
CC from CEM cells infected with the virus. The library was screened
CC with a pB1 probe corresp. to a fragment from HIV 1. The virus is
CC more cytopathic than other strains and is not inhibited by OKT4.
CC It has been deposited as CNCM I-857. The sequence can be used to
CC express proteins useful for diagnosing the presence of NDK and
CC related viruses and in vaccines against immunodeficiency diseases.
CC See also AAR09301-5.

XX SQ Sequence 119 AA;

Query Match 45.1%; Score 37; DB 11; Length 119;
Best Local Similarity 63.6%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLLEVGVEE 13
DB 109 ppavlesgte 119

RESULT 26

AAR20504

ID AAR20504 standard; Protein; 398 AA.

XX AC AAR20504;

DT 06-MAY-1992 (first entry)

XX DE Mutant Human alpha-1-antichymotrypsin with Arg(358).

XX KW ACT; unglycosylated; serine protease inhibitor; serpin.

XX OS Homo sapiens.

XX PN US5079336-A.

XX PD 07-JAN-1992.

XX PF 23-JUN-1989; 89US-0370704.

XX PR 23-JUN-1989; 89US-0370704.

XX PA (UYPE-) UNIV OF PENNSYLVANI.

XX PI Rubin H, Wang ZM;

XX DR WPI; 1992-041064/05.

XX PT New alpha-1-anti-chymotrypsin produced by recombinant DNA
PT technology - for treatment of septic shock, pancreatitis,
PT coagulation disorders, skin inflammation, etc.

XX PS Claim 5; Column 16; 15pp; English.

XX CC The amino acid sequence of wild-type ACT was derived from a cDNA
CC clone which contained the entire human ACT coding region. Mutations
CC were introduced into the coding sequence using site-directed
CC techniques. In this mutant ACT, the active site (amino acids 358-9)
CC has been mutated such that wild-type Leu at position 358 is
CC substituted by Arg. The mutant has antichymotrypsin activity and
CC anti-trypsin and antithrombin activity and can be used for treating
CC blood clotting diseases and pancreatitis.
CC See also AAQ20797 and AAR20503-R20507.

XX SQ Sequence 398 AA;

Query Match 45.1%; Score 37; DB 13; Length 398;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLELVGVVEKF 15
DB 302 illqgieeaf 312

RESULT 27

AAR20505

ID AAR20505 standard; Protein; 398 AA.

XX AC AAR20505;

DT 06-MAY-1992 (first entry)

XX DE Mutant Human alpha-1-antichymotrypsin with Met(358).

XX KW ACT; unglycosylated; serine protease inhibitor; serpin.

XX OS Homo sapiens.

XX PN US5079336-A.

XX PD 07-JAN-1992.

XX 23-JUN-1989; 89US-0370704.
 XX 23-JUN-1989; 89US-0370704.
 XX (UYPE-) UNIV OF PENNSYLVANI.
 XX Rubin H, Wang ZM;
 XX WPI; 1992-041064/05.
 XX New alpha-1-anti-chymotrypsin produced by recombinant DNA
 XX technology - for treatment of septic shock, pancreatitis,
 XX coagulation disorders, skin inflammation, etc.
 XX Claim 6; Column 16; 15pp; English.
 XX The amino acid sequence of wild-type ACT was derived from a cDNA
 XX clone which contained the entire human ACT coding region. Mutations
 XX were introduced into the coding sequence using site-directed
 XX techniques. In this mutant ACT, the active site (amino acids 358-9)
 XX have been substituted by Gly-Thr as a result of introducing a KpnI
 XX restriction site at codon 368 and 369 has resulted in the
 XX substitution of Thr for wild-type Val at amino acid position 368.
 XX The mutations to the coding sequence allow the removal of the
 XX region coding for the active site; synthetic cassettes can be
 XX inserted to create new analogues.
 XX See also AAQ20797 and AAR20503-R20507.
 XX Sequence 398 AA;
 XX
 CC Query Match 45.1%; Score 37; DB 13; Length 398;
 CC Best Local Similarity 54.5%; Pred. No. 1.7e+02;
 CC Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 CC
 OY 5 LLELVGVVEK 15
 DB 302 illqlgieeaf 312
 DB
 XX
 XX RESULT 28
 XX AAR20506
 XX ID AAR20506 standard; Protein; 398 AA.
 XX AC AAR20506;
 XX 06-MAY-1992 (first entry)
 XX Mutant Human alpha-1-antichymotrypsin.
 XX ACT; unglycosylated; serine protease inhibitor; serpin.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 XX Active-site 358..359
 XX US5079336-A.
 XX 07-JAN-1992.
 XX 23-JUN-1989; 89US-0370704.
 XX 23-JUN-1989; 89US-0370704.
 XX (UYPE-) UNIV OF PENNSYLVANI.
 XX Rubin H, Wang ZM;
 XX WPI; 1992-041064/05.
 XX New alpha-1-anti-chymotrypsin produced by recombinant DNA
 XX technology - for treatment of septic shock, pancreatitis,
 XX coagulation disorders, skin inflammation, etc.

XX Claim 7; Column 16; 15pp; English.
 XX The amino acid sequence of wild-type ACT was derived from a cDNA
 XX clone which contained the entire human ACT coding region. Mutations
 XX were introduced into the coding sequence using site-directed
 XX techniques. In this mutant ACT, amino acids Ala(349) and Ala(350)
 XX have been substituted by Gly-Thr as a result of introducing a KpnI
 XX restriction site at codon 368 and 369 has resulted in the
 XX substitution of Thr for wild-type Val at amino acid position 368.
 XX The mutations to the coding sequence allow the removal of the
 XX region coding for the active site; synthetic cassettes can be
 XX inserted to create new analogues.
 XX See also AAQ20797 and AAR20503-R20507.
 XX Sequence 398 AA;
 XX
 CC Query Match 45.1%; Score 37; DB 13; Length 398;
 CC Best Local Similarity 54.5%; Pred. No. 1.7e+02;
 CC Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 CC
 OY 5 LLELVGVVEK 15
 DB 302 illqlgieeaf 312
 DB
 XX
 XX RESULT 29
 XX AAR20507
 XX ID AAR20507 standard; Protein; 398 AA.
 XX AC AAR20507;
 XX 06-MAY-1992 (first entry)
 XX Mutant Human alpha-1-antichymotrypsin with Phe(358).
 XX ACT; unglycosylated; serine protease inhibitor; serpin.
 XX Homo sapiens.
 XX US5079336-A.
 XX 07-JAN-1992.
 XX 23-JUN-1989; 89US-0370704.
 XX 23-JUN-1989; 89US-0370704.
 XX (UYPE-) UNIV OF PENNSYLVANI.
 XX Rubin H, Wang ZM;
 XX WPI; 1992-041064/05.
 XX New alpha-1-anti-chymotrypsin produced by recombinant DNA
 XX technology - for treatment of septic shock, pancreatitis,
 XX coagulation disorders, skin inflammation, etc.
 XX Claim 8; Column 16; 15pp; English.
 XX The amino acid sequence of wild-type ACT was derived from a cDNA
 XX clone which contained the entire human ACT coding region. Mutations
 XX were introduced into the coding sequence using site-directed
 XX techniques. In this mutant ACT, the active site (amino acids 358-9)
 XX has been mutated such that wild-type Leu at position 358 is
 XX substituted by Phe. The mutant has antichymotrypsin activity and
 XX anti-trypsin and antithrombin activity and can be used for treating
 XX blood clotting diseases and pancreatitis.
 XX See also AAQ20797 and AAR20503-R20506.
 XX Sequence 398 AA;
 XX

Query Match 45.1%; Score 37; DB 13; Length 398;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEERF 15
:|:|:|:|:|
Db 302 illqlgieeaf 312

RESULT 30
AAR83104
ID AAR83104 standard; Protein; 401 AA.
XX
AC AAR83104;
XX
DT 10-JUN-1996 (first entry)
XX
DE Alpha-1-antichymotrypsin rMAS-ACT analogue.
XX
KW Alpha-1-antichymotrypsin; serine protease inhibitor; serpin;
KW chymase inhibitor; protein engineering; analogue; infection;
KW reperfusion injury; blood clot; lung inflammation; therapeutic.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 1..3
FT /label= MAS
FT /note= "N-terminal extension (replaces LCHP of
FT AAR83101)"
FT Protein 4..401
FT /note= "mature protein"
FT FT
FT W09527053-A1.
FT 12-OCT-1995.
XX
XX 29-APR-1994; 94WO-US04703.
XX
XX 31-MAR-1994; 94US-0221171.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Cooperman BS, Rubin H, Schechter N, Wang ZM;
XX WPI; 1995-366158/47.
XX
XX Analogues of human alpha-1-antichymotrypsin with increased
XX inhibitory activity - useful for treating lung inflammation etc.
XX PT also related nucleic acid, vectors and transformed cells.
XX
XX Claim 8; Fig 1A-1E; 54pp; English.
XX
XX Analogues of wild-type alpha-1-antichymotrypsin (al-ACT) (AAR83101)
XX may be produced, by recombinant DNA methods, in which amino acids
XX 356-361 of the mature protein are replaced by other residues (e.g.
XX see AAR83102-03), and optionally contain a modified N-terminal
XX extension (see also AAR83105-06). The N-terminal extensions M-A-S
XX or A-S on the wild-type and analogue proteins enable direct
XX expression of stable monomers. The wild-type protein and analogues
XX are used as chymase inhibitors and in the treatment and prevention
XX of blood clots, reperfusion injury and lung inflammation, the
XX latter caused by acid inhalation (from stomach contents or smoke)
XX or infection by a Gram-negative bacterium (e.g. Pseudomonas or
XX Escherichia). The analogues are prepared by protein engineering
XX techniques, and are 4-fold more efficient at inhibiting chymase
XX than the wild-type protein.
XX Sequence 401 AA;

Query Match 45.1%; Score 37; DB 16; Length 401;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEERF 15
:|:|:|:|:|
Db 305 illqlgieeaf 315

RESULT 31
AAR83105
ID AAR83105 standard; Protein; 401 AA.
XX
AC AAR83105;
XX
DT 10-JUN-1996 (first entry)
XX
DE Alpha-1-antichymotrypsin rMAS-ACT analogue.
XX
KW Alpha-1-antichymotrypsin; serine protease inhibitor; serpin;
KW chymase inhibitor; protein engineering; analogue; infection;
KW reperfusion injury; blood clot; lung inflammation; therapeutic.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 1..3
FT /label= MAS
FT /note= "N-terminal extension (replaces LCHP of
FT AAR83102)"
FT Protein 4..401
FT /note= "mature protein"
FT FT
FT Misc-difference 361
FT /label= Trp, Ala, Asn, Asp, Cys, Gln, Glu, Gly,
FT His, Ile, Lys, Met, Phe, Pro, Ser, Thr,
FT Tyr, Val
FT /note= "wild-type Leu-358 substitution"
FT FT
FT Misc-difference 361
FT /label= L358X
FT /note= "Trp (claim 6) (see AAT00428)"
XX
XX W09527053-A1.
XX
XX 12-OCT-1995.
XX
XX 29-APR-1994; 94WO-US04703.
XX
XX 31-MAR-1994; 94US-0221171.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Cooperman BS, Rubin H, Schechter N, Wang ZM;
XX WPI; 1995-366158/47.
XX
XX Analogues of human alpha-1-antichymotrypsin with increased
XX inhibitory activity - useful for treating lung inflammation etc.
XX PT also related nucleic acid, vectors and transformed cells.
XX
XX Claim 8; Fig 1A-1E; 54pp; English.
XX
XX Analogues of wild-type alpha-1-antichymotrypsin (al-ACT) (AAR83101)
XX may be produced, by recombinant DNA methods, in which amino acids
XX 356-361 of the mature protein are replaced by other residues (e.g.
XX see AAR83102-03), and optionally contain a modified N-terminal
XX extension (see also AAR83104-06). The N-terminal extensions M-A-S
XX or A-S on the wild-type and analogue proteins enable direct
XX expression of stable monomers. The wild-type protein and analogues
XX are used as chymase inhibitors and in the treatment and prevention
XX of blood clots, reperfusion injury and lung inflammation, the
XX latter caused by acid inhalation (from stomach contents or smoke)
XX or infection by a Gram-negative bacterium (e.g. Pseudomonas or
XX Escherichia). The analogues are prepared by protein engineering
XX techniques, and are 4-fold more efficient at inhibiting chymase
XX than the wild-type protein.
XX Sequence 401 AA;

CC techniques, and are 4-fold more efficient at inhibiting chymase
 CC than the wild-type protein.
 XX
 SQ Sequence 401 AA;

Query Match Similarity 45.1%; Score 37; DB 16; Length 401;
 Best Local Similarity 54.5%; Pred. No. 1; 7e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 LLEVGVEEFK 15
 DB 305 illqlgileaf 315

RESULT 32

ID AAR83106 standard; Protein; 401 AA.

XX AAR83106;

DE 10-JUN-1996 (first entry)

DE Alpha-1-antichymotrypsin rMAS-ACT analogue.

XX Alpha-1-antichymotrypsin; serine protease inhibitor; serpin;
 KW chymase inhibitor; protein engineering; analogue; infection;
 KW reperfusion injury; blood clot; lung inflammation; therapeutic.
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 1..3
 FT /label= MAS
 FT /note= "N-terminal extension (replaces LCHP of
 FT AAR83103)"

FT Protein 4..401

FT /note= "mature protein"

FT Misc-difference 359..364

FT /label= IPXSIP

FT /note= "replaces wild-type TLLSAL"

FT Misc-difference 260..261
 FT /label= Met, Trp, Ala, Asn, Asp, Cys, Gln, Glu,
 FT Gly, His, Ile, Lys, Phe, Pro, Ser, Thr,
 FT Tyr, Val
 FT /note= "pref. Met (claim 3)"

XX WO9527053-A1.

XX 12-OCT-1995.

XX 29-APR-1994; 94WO-US04703.

XX 31-MAR-1994; 94US-0221171.

XX (UYPE-) UNIV PENNSYLVANIA.

XX Cooperman BS, Rubin H, Schechter N, Wang ZM;

XX WPI; 1995-366158/47.

XX Analogues of human alpha-1-anti-chymotrypsin with increased
 PT inhibitory activity, effective in lung inflammation etc.
 PT also related nucleic acid, vectors and transformed cells.

XX Claim 8; Fig 1A-1E; 54pp; English.

XX Analogues of wild-type alpha-1-antichymotrypsin (al-ACT) (AAR83101)
 CC may be produced, by recombinant DNA methods, in which amino acids
 CC are replaced by other amino acids, e.g., by glutamic acid (e.g.
 CC see AAR83102-03) and optionally contain a modified N-terminal
 CC extension (see also AAR83104-05). The N-terminal extensions M-A-S
 CC or A-S on the wild-type and analogue proteins enable direct

CC expression of stable monomers. The wild-type protein and analogues
 CC are used as chymase inhibitors and in the treatment and prevention
 CC of blood clots, reperfusion injury and lung inflammation, the
 CC latter caused by acid inhalation (from stomach contents or smoke)
 CC or infection by a Gram-negative bacterium (e.g. Pseudomonas or
 CC Escherichia). The analogues are prepared by protein engineering
 CC techniques and are 4-fold more efficient at inhibiting chymase
 CC than the wild-type protein.
 XX Sequence 401 AA;

Query Match 45.1%; Score 37; DB 16; Length 401;
 Best Local Similarity 54.5%; Pred. No. 1; 7e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 LLEVGVEEFK 15
 DB 305 illqlgileaf 315

RESULT 33

ID AAR44435 standard; Protein; 402 AA.

XX AAR44435;

XX 08-JUN-1994 (first entry)

DE Alpha-antichymotrypsin.

KW Alpha-antichymotrypsin; recombinant; septic shock; coagulation;
 KW disorders; liver disease; skin; inflammation.

XX Homo sapiens.

XX Key

FT Peptide Location/Qualifiers

FT Active-site 4..398

FT /label= mat_protein

FT /note= 362..363

XX US5266465-A.

XX 30-NOV-1993.

XX 23-JUN-1989; 89US-0370704.

XX 23-JUN-1989; 89US-0370704.

XX 24-JUL-1991; 91US-0735322.

XX (UYPE-) UNIV PENNSYLVANIA.

XX Cooperman BS, Rubin H, Schechter N, Wang ZM;

XX WPI; 1993-305337/49.

XX P-PSDB; AAR44435.

XX Prodn. of analogues of alpha-1-anti-chymotrypsin useful in
 PT treating septic shock, skin inflammation etc.; - by recombinant
 PT methods in which most cells transformed with DNA encoding
 PT analogue are cultured.

XX Disclosure; Fig 1a-1e; 15pp; English.

XX The sequence is of an alpha-1-antichymotrypsin. The protein
 CC can be used to treat septic shock, blood clotting disorders,
 CC pancreatitis and liver diseases. The gene was manufactured by site
 CC directed mutagenesis using a commercially available kit.

XX Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;

Best Local Similarity 54.5%; Pred. No. 1.7e+02; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEERF 15
:|:|:|:|
Db 306 illqlgleeaf 316

RESULT 34
AAR42737
ID AAR42737 standard; Protein; 402 AA.
XX
AC AAR42737;
DT
DT 05-MAY-1994 (first entry)
XX
DE Alpha-1-antichymotrypsin (Leu358Arg).
XX
KW Analogue; alpha-1-antichymotrypsin; ACT; septic shock;
KW pancreatitis; coagulation disorder; liver disease; enzyme;
KW skin; inflammation; anti-thrombin; anti-trypsin.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 5..402
FT /label= mat_protein
XX
PN US5252725-A.
XX
PD 12-OCT-1993.
XX
PF 23-JUN-1989; 89US-0370704.
XX
PR 23-JUN-1989; 89US-0370704.
PR 24-JUL-1991; 91US-0735335.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
PI Ruben H, Wang ZM;
XX
XX WPI; 1993-336158/42.
XX
XX Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -
PT used for treating e.g. septic shock, pancreatitis, coagulation
PT disorders, microbial disease and skin inflammation
XX
PS Claim 1; Fig 1; 15pp; English.

XX An isolated nucleic acid sequence encoding an analogue of ACT is
CC claimed in which (i) Leu358 is substd. with Arg, Phe or Met or (ii)
CC AlaAla349-350 are substd. with GlyThr, ValArg368-369 are substd.
CC with ThrArg and Leu358 is opt. substd. with Phe.
CC ACT analogues having the Met or Phe substd. have anti-chymotrypsin
CC activity and can be used for the treatment of septic shock,
CC pancreatitis, coagulation disorders as in liver diseases, certain
CC diseases caused by microbes that penetrate the skin by elaborating
CC a microbial chymotrypsin-like enzyme and skin inflammation.
CC ACT analogues having the Arg substd. have anti-thrombin and
CC anti-trypsin activity and can be used for treating blood clotting
CC diseases and pancreatitis.
XX
SQ Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEERF 15
:|:|:|:|
Db 306 illqlgleeaf 316

RESULT 35
AAR42738
ID AAR42738 standard; Protein; 402 AA.
XX
AC AAR42738;
DT
DT 05-MAY-1994 (first entry)
XX
DE Alpha-1-antichymotrypsin (Leu358Phe).
XX
KW Analogue; alpha-1-antichymotrypsin; ACT; septic shock;
KW pancreatitis; coagulation disorder; liver disease; enzyme;
KW skin; inflammation; anti-thrombin; anti-trypsin.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 5..402
FT /label= mat_protein
XX
PN US5252725-A.
XX
PD 12-OCT-1993.
XX
PF 23-JUN-1989; 89US-0370704.
XX
PR 23-JUN-1989; 89US-0370704.
PR 24-JUL-1991; 91US-0735335.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
PI Ruben H, Wang ZM;
XX
XX WPI; 1993-336158/42.
XX
XX Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -
PT used for treating e.g. septic shock, pancreatitis, coagulation
PT disorders, microbial disease and skin inflammation
XX
PS Claim 4; Fig 1; 15pp; English.

XX An isolated nucleic acid sequence encoding an analogue of ACT is
CC claimed in which (i) Leu358 is substd. with Arg, Phe or Met or (ii)
CC AlaAla349-350 are substd. with GlyThr, ValArg368-369 are substd.
CC with ThrArg and Leu358 is opt. substd. with Phe.
CC ACT analogues having the Met or Phe substd. have anti-chymotrypsin
CC activity and can be used for the treatment of septic shock,
CC pancreatitis, coagulation disorders as in liver diseases, certain
CC diseases caused by microbes that penetrate the skin by elaborating
CC a microbial chymotrypsin-like enzyme and skin inflammation.
CC ACT analogues having the Arg substd. have anti-thrombin and
CC anti-trypsin activity and can be used for treating blood clotting
CC diseases and pancreatitis.
XX
SQ Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEERF 15
:|:|:|:|
Db 306 illqlgleeaf 316

RESULT 36
AAR42739
ID AAR42739 standard; Protein; 402 AA.
XX
AC AAR42739;
DT
DT 05-MAY-1994 (first entry)
XX
DE Alpha-1-antichymotrypsin (Leu358Arg).
XX
KW Analogue; alpha-1-antichymotrypsin; ACT; septic shock;
KW pancreatitis; coagulation disorder; liver disease; enzyme;
KW skin; inflammation; anti-thrombin; anti-trypsin.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 5..402
FT /label= mat_protein
XX
PN US5252725-A.
XX
PD 12-OCT-1993.
XX
PF 23-JUN-1989; 89US-0370704.
XX
PR 23-JUN-1989; 89US-0370704.
PR 24-JUL-1991; 91US-0735335.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
PI Ruben H, Wang ZM;
XX
XX WPI; 1993-336158/42.
XX
XX Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -
PT used for treating e.g. septic shock, pancreatitis, coagulation
PT disorders, microbial disease and skin inflammation
XX
PS Claim 4; Fig 1; 15pp; English.

05-MAY-1994 (first entry)
Alpha-1-antichymotrypsin (Leu358Met).
Analogue; alpha-1-antichymotrypsin; ACT; septic shock;
pancreatitis; coagulation disorder; liver disease; enzyme;
skin; inflammation; anti-thrombin; anti-trypsin.

Homo sapiens.

Key Location/Qualifiers
FT 5..402
FT /label= mat_protein

US5252725-A.

12-OCT-1993.

23-JUN-1989; 89US-0370704.

23-JUN-1989; 89US-0370704.

24-JUL-1991; 91US-0735335.

(UYPE-) UNIV PENNSYLVANIA.

Rubin H, Wang ZM;

WPI; 1993-336158/42.

Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -
used for treating e.g. septic shock, pancreatitis, coagulation
disorders, microbial disease and skin inflammation

Claim 2; Fig 1; 15pp; English.

An isolated nucleic acid sequence encoding an analogue of ACT is
claimed in which (i) Leu358 is substid with Arg, Phe or Met or (ii)
Ala349-350 are substid with Glythr, Valarg368-369 are substid.
ACT analogues having the Met or Phe substn. have anti-chymotrypsin
activity and can be used for the treatment of septic shock, certain
pancreatitis, coagulation disorders as in liver diseases, certain
diseases caused by microbes that penetrate the skin by elaborating
a microbial chymotrypsin-like enzyme and skin inflammation and
ACT analogues having the Arg substn. have anti-thrombin and
anti-trypsin activity and can be used for treating blood clotting
diseases and pancreatitis.

Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;
Best Local Similarity 54.5%; Pred. NO. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 LLELVGVVEERF 15

Db 306 illqlgileaf 316

RESULT 37

AAR42740

AAR42740 standard; Protein; 402 AA.

AC AAR42740;

05-MAY-1994 (first entry)

Alpha-1-antichymotrypsin analogue.

Analogue; alpha-1-antichymotrypsin; ACT; septic shock;
pancreatitis; coagulation disorder; liver disease; enzyme;
skin; inflammation; anti-thrombin; anti-trypsin.

XX

OS Homo sapiens.

Key Location/Qualifiers

FT 5..402

FT /label= mat_protein

FT Misc-difference 362

FT /note= "Leu opt. substd. with Phe"

FT Misc-difference 355

FT /note= "AlaAla(349-350) substd. with GlyThr"

FT Misc-difference 372..373

FT /note= "ValArg(368-369) substd. with ThrArg"

XX US5252725-A.

XX 12-OCT-1993.

XX 23-JUN-1989; 89US-0370704.

XX 23-JUN-1989; 89US-0370704.

XX 24-JUL-1991; 91US-0735335.

XX (UYPE-) UNIV PENNSYLVANIA.

XX Rubin H, Wang ZM;

XX WPI; 1993-336158/42.

Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -
used for treating e.g. septic shock, pancreatitis, coagulation
disorders, microbial disease and skin inflammation

Claim 3-4; Fig 1; 15pp; English.

An isolated nucleic acid sequence encoding an analogue of ACT is
claimed in which (i) Leu358 is substid with Arg, Phe or Met or (ii)
Ala349-350 are substid with Glythr, Valarg368-369 are substd.
ACT analogues having the Met or Phe substn. have anti-chymotrypsin
activity and can be used for the treatment of septic shock, certain
pancreatitis, coagulation disorders as in liver diseases, certain
diseases caused by microbes that penetrate the skin by elaborating
a microbial chymotrypsin-like enzyme and skin inflammation and
ACT analogues having the Arg substn. have anti-thrombin and
anti-trypsin activity and can be used for treating blood clotting
diseases and pancreatitis.

Sequence 402 AA;

Query Match 45.1%; Score 37; DB 14; Length 402;
Best Local Similarity 54.5%; Pred. NO. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 LLELVGVVEERF 15

Db 306 illqlgileaf 316

RESULT 38

AAR83101

AAR83101 standard; Protein; 402 AA.

AC AAR83101;

10-JUN-1996 (first entry)

Wild-type alpha-1-antichymotrypsin.

Alpha-1-antichymotrypsin; serine protease inhibitor; serpin;
chymase inhibitor; protein engineering; analogue; infection;
reperfusion injury; blood clot; lung inflammation; therapeutic.

```

OS Homo sapiens.
XX
PH KEY Location/Qualifiers
FT Misc-difference 1..4 /label= LCHP
FT /note= "N-terminal, opt. replaced with N-terminal
FT extension (M)AS (see AAR83105)"
FT
FT Protein 5..402 /note= "mature protein"
FT /note= "mature RACT-L358 analogue"
FT Misc-difference 362
FT /label= Trp, Ala, Asn, Asp, Cys, Gln, Glu, Gly,
FT His, Ile, Lys, Met, Phe, Pro, Ser, Thr,
FT Tyr, Val
XX
PN WO9527053-A1.
XX
XX 12-OCT-1995.
XX
XX 29-APR-1994; 94WO-US04703.
XX
XX 31-MAR-1994; 94US-0221171.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Cooperman BS, Rubin H, Schechter N, Wang ZM;
XX WPI; 1995-366158/47.
XX N-PSDB; AAT00427.
XX
XX Analogues of human alpha-1-antichymotrypsin with increased
XX inhibitory activity - useful for treating lung inflammation etc.
XX PT also related nucleic acid, vectors and transformed cells.
XX
XX Claim 17; Fig 1a-1e; 54pp; English.
XX
XX Wild-type alpha-1-antichymotrypsin (a1-ACT) may be produced by
XX recombinant DNA methods. Analogues of a1-ACT (AAR83102-06) may be
XX produced in which amino acids 356-361 of the mature protein are
XX replaced by other residues, and optionally contain a modified
XX N-terminal extension. The wild-type protein and analogues
XX are used as chymase inhibitors and in the treatment and prevention
XX of blood clots, reperfusion injury and lung inflammation, the latter
XX caused by acid inhalation (from stomach contents or smoke) or
XX infection by a Gram-negative bacterium (e.g. Pseudomonas or
XX Escherichia). The analogues are prepared by protein engineering
XX techniques, and are 4-fold more efficient at inhibiting chymase than
XX the wild-type protein.
XX
XX Sequence 402 AA;

Query Match 45.1%; Score 37; DB 16; Length 402;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEERF 15
Db 306 illqlgieeaf 316
:|:|:|:|

RESULT 39
AAR83102
ID AAR83102 standard; Protein; 402 AA.
XX
XX AAR83102;
XX
XX 10-JUN-1996 (first entry)
XX
XX Alpha-1-antichymotrypsin analogue.
XX
XX Alpha-1-antichymotrypsin analogue; serine protease inhibitor;
XX serpin; chymase inhibitor; protein engineering; infection;
XX reperfusion injury; blood clot; lung inflammation; therapeutic.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers

```

```

FT Misc-difference 1..4 /label= LCHP
FT /note= "N-terminal, opt. replaced with N-terminal
FT extension (M)AS (see AAR83105)"
FT
FT Protein 5..402 /note= "mature RACT-L358 analogue"
FT Misc-difference 362
FT /label= Trp, Ala, Asn, Asp, Cys, Gln, Glu, Gly,
FT His, Ile, Lys, Met, Phe, Pro, Ser, Thr,
FT Tyr, Val
XX
PN WO9527053-A1.
XX
XX 12-OCT-1995.
XX
XX 29-APR-1994; 94WO-US04703.
XX
XX 31-MAR-1994; 94US-0221171.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
XX
XX Cooperman BS, Rubin H, Schechter N, Wang ZM;
XX WPI; 1995-366158/47.
XX
XX Analogues of human alpha-1-antichymotrypsin with increased
XX inhibitory activity - useful for treating lung inflammation etc.
XX PT also related nucleic acid, vectors and transformed cells.
XX
XX Claim 1; Fig 1a-1e; 54pp; English.
XX
XX The sequence represent an analogue of wild-type alpha-1-
XX antichymotrypsin (a1-ACT, AAR83101) having an alteration at Ieu358
XX as shown above. Other analogues may be produced in which the
XX 356-361 AA sequence of the mature protein is replaced by
XX another sequence (see AAR83103), optionally with a modified N-terminal
XX extension (see AAR83104-06). The wild-type protein and analogues are
XX used as chymase inhibitors and in the treatment and prevention of
XX blood clots, reperfusion injury and lung inflammation, the latter
XX caused by acid inhalation (from stomach contents or smoke) or
XX infection by a Gram-negative bacterium (e.g. Pseudomonas or
XX Escherichia). The analogues are prepared by protein engineering
XX techniques, and are 4-fold more efficient at inhibiting chymase
XX than the wild-type protein.
XX
XX Sequence 402 AA;

Query Match 45.1%; Score 37; DB 16; Length 402;
Best Local Similarity 54.5%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEERF 15
Db 306 illqlgieeaf 316
:|:|:|:|

RESULT 40
AAR83103
ID AAR83103 standard; Protein; 402 AA.
XX
XX AAR83103;
XX
XX 10-JUN-1996 (first entry)
XX
XX Alpha-1-antichymotrypsin analogue.
XX
XX Alpha-1-antichymotrypsin; serine protease inhibitor; serpin;
XX chymase inhibitor; protein engineering; analogue; infection;

```

KW reperfusion injury; blood clot; lung inflammation; therapeutic.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1.4
 FT /label= ICHP
 FT /note= "terminal, opt. replaced with N-terminal
 FT extension (N)AS (see AAR83106)"
 FT Protein 5..402
 FT /note= "mature protein"
 FT Misc-difference 360..365
 FT /label= IPXSTP
 FT /note= "replaces wild-type TLISAL"
 FT Misc-difference 360..365
 FT /label= Met, Trp, Ala, Asn, Asp, Cys, Gln, Glu,
 FT Gly, His, Ile, Lys, Phe, Pro, Ser, Thr,
 FT Tyr, Val
 FT /note= "pref. Met (claim 3)"
 XX WO9527053-A.
 XX 12-OCT-1995.
 PD XX
 XX 29-APR-1994; 94WO-US04703.
 XX 31-MAR-1994; 94US-0221171.
 XX (UYPE-) UNIV PENNSYLVANIA.
 XX Cooperman BS, Rubin H, Schechter N, Wang ZM;
 XX WPI; 1995-366158/47.
 XX
 DR Analogs of human alpha-1-anti-chymotrypsin with increased
 PT inhibitory activity useful for treating lung inflammation etc.
 PT also related nucleic acid vectors and transformed cells.
 XX Claim 2; Fig 1A-1E; 54pp; English.
 XX The sequence represent an analog of wild-type alpha-1-
 CC anti-chymotrypsin (AAR83102). The wild-type may be
 CC produced (see AAR83102) optionally with modified N-terminal
 CC extensions (see AAR83104-06). The wild-type protein and analogues
 CC are used as chymase inhibitors and in the treatment and prevention
 CC of blood clots, reperfusion injury and lung inflammation, the latter
 CC caused by acid inhalation (from stomach contents or smoke) or
 CC infection by a gram-negative bacterium (e.g. Pseudomonas or
 CC Klebsiella). The analogues are prepared by recombinant
 CC techniques and are 4-fold more efficient at inhibiting chymase than
 CC the wild-type protein.
 XX Sequence 402 AA;

Query Match 45.1%; Score 37; DB 16; Length 402;
 Best Local Similarity 54.5%; Pred. No. 1.7e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 5 LLEVGVEEFK 15
 Db 306 illqlgieef 316
 RESULT 41
 AAR67259
 ID AAR67259 standard; Protein; 402 AA.
 XX
 XX AAR67259;
 XX 03-JUL-1995 (first entry)
 DT Alpha-1-antichymotrypsin.
 DE

XX Alpha-1-antichymotrypsin; ACT; antithrombin; antitrypsin;
 KW chymotrypsin-inhibitor; protease-inhibitor; blood clotting;
 KW pancreatitis.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FT Reptide 1..4
 FT /note= "amino acids at positions 1-4 appear
 FT in the precursor of the mature protein"
 FT US5367064-A.
 XX 22-NOV-1994.
 XX 23-JUN-1989; 89US-0370704.
 XX 23-JUN-1989; 89US-0370704.
 XX 24-JUL-1991; 91US-0735335.
 XX 15-JAN-1993; 93US-0005908.
 XX (UYPE-) UNIV PENNSYLVANIA.
 XX Cooperman BS, Rubin H, Schechter N, Wang ZM;
 XX WPI; 1995-005897/01.
 XX N-PSDB; AAQ/5286.
 XX Nucleic acid encoding alpha-1-anti-chymotrypsin - for production
 PT of recombinant ACT and analogues having anti-trypsin,
 PT antithrombin and anti-chymotrypsin activity, useful in the
 PT treatment of blood clotting and pancreatitis
 XX Disclosure; Fig. 1a-1d; 16pp; English.
 XX A human liver cDNA library in phage lambda-gt11 was screened with
 CC polyclonal antisera raised against Cl esterase-inhibitor. The DNA
 CC sequence of the insert of a positive clone contained the entire
 CC coding region of human alpha-antichymotrypsin.
 XX Sequence 402 AA;
 QY 5 LLEVGVEEFK 15
 Db 306 illqlgieef 316
 Query Match 45.1%; Score 37; DB 16; Length 402;
 Best Local Similarity 54.5%; Pred. No. 1.7e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 5 LLEVGVEEFK 15
 Db 306 illqlgieef 316
 RESULT 42
 AAR20503
 ID AAR20503 standard; Protein; 476 AA.
 XX
 XX AAR20503;
 XX 06-MAY-1992 (first entry)
 DT Human alpha-1-antichymotrypsin.
 DE
 XX ACT; unglycosylated; serine protease inhibitor; serpin.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FT Protein 5..402
 FT /label= ACT
 FT Active-site 362..363
 XX US5079336-A.
 PN

XX PD 07-JAN-1992.
 XX PF 23-JUN-1989; 89US-0370704.
 XX PR 23-JUN-1989; 89US-0370704.
 XX PR 23-JUN-1989; 89US-0370704.
 XX PA (UYPE-) UNIV OF PENNSYLVANIA.
 XX PI Rubin H, Wang ZM;
 XX PI WPI; 1992-041064/05.
 XX DR N-PSDB; AAQ20797.
 XX PT New alpha-1-anti-chymotrypsin produced by recombinant DNA
 XX PT technology - for treatment of septic shock, pancreatitis,
 XX PT coagulation disorders, skin inflammation, etc.
 XX PS Claim 3; Fig 1; 15pp; English.
 XX CC This amino acid sequence was derived from a cDNA clone which
 XX CC contained the entire human ACT coding region. The C-terminal
 XX CC sequence is in agreement with Hill et al., Nature 311: 175-177,
 XX CC (1984) and the remainder of the sequence is in agreement with
 XX CC Chandra et al., Biochemistry 22: 5055-5060, (1983) except for the
 XX CC 15 amino acids from position 77 to 91 and the 6 amino acids from 98
 XX CC to 103. The differences are due to 3 insertions and 3 deletions of
 XX CC single bases within the Chandra sequence. The sequence reported
 XX CC here shows a Pro at position 44 rather than a leucine, a leucine
 XX CC at position 174, rather than a Pro, an Ala for val at 336 and a Leu
 XX CC for Ser at 338. The latter two substitutions have been reported by
 XX CC Hill et al.. A second form of mature protein that includes two
 XX CC additional amino acids, His-Pro, at the terminus has also been
 XX CC found. The "X" in the sequence corresponds to a nonsense (i.e.
 XX CC STOP) codon. The specification includes the translated sequence of
 XX CC amino acids beyond the termination codon (403-476).
 XX CC See also AAR20504-7.
 XX Sequence 476 AA;

Query Match 45.1%; Score 37; DB 13; Length 476;
 Best Local Similarity 54.5%; Pred. No. 2.1e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEEKF 15
 :||:|:|:|
 Db 306 illqlgieeaf 316

RESULT 43
 AAR42736
 ID AAR42736 standard; Protein; 476 AA.

XX AC AAR42736;
 XX DT 05-MAY-1994 (first entry)
 XX DE Alpha-1-antichymotrypsin.

XX KW Analogue; alpha-1-antichymotrypsin; ACT; septic shock;
 XX KW pancreatitis; coagulation disorder; liver disease; enzyme;
 XX KW skin; inflammation; anti-thrombin; anti-trypsin.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT Misc-difference 403
 FT /note= "corresp. to stop codon"

XX PN US5252725-A.
 XX XX 12-OCT-1993.

XX 23-JUN-1989; 89US-0370704.
 XX 23-JUN-1989; 89US-0370704.
 XX 24-JUL-1991; 91US-0735335.
 XX (UYPE-) UNIV PENNSYLVANIA.
 XX PI Rubin H, Wang ZM;
 XX PI WPI; 1993-336158/42.
 XX DR N-PSDB; AAQ49263.
 XX PT Nucleic acid encoding human alpha-1-anti-chymotrypsin analogues -
 XX PT used for treating e.g. septic shock, pancreatitis, coagulation
 XX PT disorders, microbial disease and skin inflammation
 XX PS Disclosure; Fig 1; 15pp; English.
 XX CC An isolated nucleic acid sequence encoding an analogue of ACT is
 XX CC claimed in which (i) Leu358 is substd. with Arg, Phe or Met or (ii)
 XX CC AlaAla349-350 are substd. with GlyThr, ValArg368-369 are substd.
 XX CC with ThrArg and Leu358 is opt. substd. with Phe.
 XX CC ACT analogues having the Met or Phe substd. have anti-chymotrypsin
 XX CC activity and can be used for the treatment of septic shock,
 XX CC pancreatitis, coagulation disorders as in liver diseases, certain
 XX CC diseases caused by microbes that penetrate the skin by elaborating
 XX CC a microbial chymotrypsin-like enzyme and skin inflammation.
 XX CC ACT analogues having the Arg substd. have anti-thrombin and
 XX CC anti-trypsin activity and can be used for treating blood clotting
 XX CC diseases and pancreatitis.
 XX Sequence 476 AA;

Query Match 45.1%; Score 37; DB 14; Length 476;
 Best Local Similarity 54.5%; Pred. No. 2.1e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEEKF 15
 :||:|:|:|
 Db 306 illqlgieeaf 316

RESULT 44
 AAR82250
 ID AAR82250 standard; Protein; 476 AA.

XX AC AAR82250;

XX DT 28-JAN-1996 (first entry)
 XX DE Mature human wild type alpha-1-antichymotrypsin.

XX KW Antichymotrypsin; ACT; serine protease inhibitor;
 XX KW neutrophil elastase.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT Misc-difference 403
 FT /label= site of stop codon in AAT03876
 FT Misc-difference 356..361
 FT /label= AAR82601
 FT /note= "substd. with AAR82602 & resultant
 FT analogue claimed"

FT Misc-difference 349..350

FT /label= G-T

FT /note= "claimed analogue"

FT Misc-difference 368..369

FT /label= T-R

FT /note= "claimed analogue"

FT Misc-difference 356..357

```

FT  /label= either or both substd. with p
FT  /note= "Claimed analogue(s)"
FT  Misc-difference 359
FT  /label= substd. with p
FT  /note= "Claimed analogue"
FT  Misc-difference 360..361
FT  /label= either or both substd. with p
FT  /note= "Claimed analogue(s)"
FT  Misc-difference 359
FT  /label= M,I,V,A,D,T,E
FT  /note= "Claimed analogue(s)"
FT  W09527055-A.
PN  12-OCT-1995.
XX  29-APR-1994; 94WO-US04735.
XX  31-MAR-1994; 94US-0221078.
XX  (UYPE-) UNIV PENNSYLVANIA.
XX  Cooperman BS, Rubin H, Schechter N, Wang ZM;
XX  WPI; 1995-358630/46.
XX  N-PSDB; AAT03876.
XX  Analogue of human alpha-1-antichymotrypsin - acts as an efficient
XX  inhibitor of human neutrophil elastase, for use in treating inflammation
XX  Disclosure; Fig 1; 51pp; English.
XX  A human liver cDNA library was screened with polyclonal antisera
XX  raised against Cl esterase inhibitor. Positive clones were
XX  sequenced. The DNA sequence and the derived AA sequence of the Eco
XX  contained the entire coding region of the mature human alpha-1-
XX  antichymotrypsin (alpha AC), as depicted in AAT03876. The construct
XX  also included a 21 nt extension of the 5'-end encoding 7 AAs,
XX  comprising the sequence in AAR2603. The mature protein has an Mr
XX  45,031. AAs corresp. to AAR2601 at AA posns. 356-361 were substd.
XX  alpha-1-prot-AAR2601 was selected from this corresp. segment of
XX  proteinase inhibitor from the p3 through the p3' posns. The
XX  resulting analogue is claimed. Other analogues are also claimed
XX  (see AAR82250 FT). The analogues have neutrophil elastase
XX  inhibiting activity.
XX  Sequence 476 AA;
XX  Query Match 45.1%; Score 37; DB 16; Length 476;
XX  Best Local Similarity 54.5%; Pred. No. 2.le+02;
XX  Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Oy 5 LLELVGVKKF 15
Db 306 illqlgieaf 316
RESULT 45
AAG31539
XX AAG31539 standard; Protein; 1410 AA.
XX AAG31539;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 37892.
XX protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.

```

```

XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 99US-0121825.
XX 99US-0123180.
XX 99US-0123548.
XX 99US-0125788.
XX 99US-0126264.
XX 99US-0127782.
XX 99US-0128234.
XX 99US-0128714.
XX 99US-0128845.
XX 99US-0130077.
XX 99US-0130449.
XX 99US-0130801.
XX 99US-0130931.
XX 99US-0131449.
XX 99US-0132048.
XX 99US-0132407.
XX 99US-0132484.
XX 99US-0132485.
XX 99US-0132487.
XX 99US-0132863.
XX 99US-0134256.
XX 99US-0134218.
XX 99US-0134219.
XX 99US-0134221.
XX 99US-0134768.
XX 99US-0134941.
XX 99US-0135134.
XX 99US-0135353.
XX 99US-0135629.
XX 99US-0136321.
XX 99US-0136921.
XX 99US-0136782.
XX 99US-0137222.
XX 99US-0137528.
XX 99US-0137502.
XX 99US-0137724.
XX 99US-0138184.
XX 99US-0138540.
XX 99US-0138847.
XX 99US-0139119.
XX 99US-0139452.
XX 99US-0139453.
XX 99US-0139454.
XX 99US-0139455.
XX 99US-0139456.
XX 99US-0139457.
XX 99US-0139458.
XX 99US-0139459.
XX 99US-0139460.
XX 99US-0139461.
XX 99US-0139462.
XX 99US-0139463.
XX 99US-0139750.
XX 99US-0139763.
XX 99US-0139817.
XX 99US-0139833.
XX 99US-0140354.
XX 99US-0140695.
XX 99US-0140823.

```

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 45.1%; Score 37; DB 21; Length 1410;
Best Local Similarity 63.6%; Pred. No. 6.7e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 PPLLEVGVEE 13
Db 668 ppliekfgrvee 678
|||: : ||||

Search completed: June 28, 2001, 11:54:12
Job time: 73 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 37.02 seconds
(without alignments)
8.706 Million cell updates/sec

Title: US-09-439-313-554
Perfect score: 82
Sequence: 1 YVPPLLLEVGVEKFM 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 65 summaries

Database : Issued Patents_AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	48.8	51	1	US-08-346-849-15
2	40	48.8	51	2	US-08-293-284A-15
3	39	47.6	1544	4	US-09-413-814-46
4	36	43.9	415	1	US-08-381-936-2
5	36	43.9	415	3	US-08-943-374-2
6	35.5	43.3	525	2	US-08-749-902-7
7	35.5	43.3	525	2	US-08-749-902-8
8	35	42.7	32	3	US-08-938-548B-7
9	35	42.7	80	3	US-08-554-840-13
10	35	42.7	123	3	US-08-938-548B-10
11	35	42.7	130	3	US-08-938-548B-6
12	35	42.7	424	4	US-08-876-885-26
13	35	42.7	486	2	US-08-432-016-6
14	35	42.7	486	2	US-08-684-594-6
15	35	42.7	2938	5	PCT-US94-00198-3
16	34	41.5	55	4	US-09-042-012-15
17	34	41.5	63	4	US-09-042-012-9
18	34	41.5	63	4	US-09-042-012-11
19	34	41.5	64	4	US-09-042-012-7
20	34	41.5	64	4	US-09-305-086-1
21	34	41.5	71	4	US-09-042-012-17
22	34	41.5	89	3	US-08-946-026-1
23	34	41.5	260	4	US-08-081-929-10
24	34	41.5	271	2	US-08-937-972-6
25	34	41.5	315	1	US-08-129-129-8
26	34	41.5	321	2	US-08-937-972-3
27	34	41.5	494	3	US-08-993-260-3

28	34	41.5	858	3	US-08-946-026-3	Sequence 3, Appli
29	34	41.5	1311	2	US-08-540-406-4	Sequence 4, Appli
30	34	41.5	1311	3	US-08-656-055-4	Sequence 4, Appli
31	34	41.5	1311	4	US-08-954-668-4	Sequence 4, Appli
32	34	41.5	1311	5	PCT-US95-13233-4	Sequence 4, Appli
33	33.5	40.9	379	4	US-09-413-304-15	Sequence 15, Appli
34	33	40.2	125	4	US-08-905-223-320	Sequence 320, App
35	33	40.2	328	2	US-08-977-767-1	Sequence 1, Appli
36	33	40.2	340	2	US-08-974-546-5	Sequence 5, Appli
37	33	40.2	366	3	US-08-987-904A-2	Sequence 2, Appli
38	33	40.2	366	3	US-08-987-904A-4	Sequence 4, Appli
39	33	40.2	395	3	US-09-080-044-5	Sequence 5, Appli
40	33	40.2	459	4	US-09-080-983-5	Sequence 9, Appli
41	33	40.2	906	3	US-08-609-230A-9	Sequence 9, Appli
42	33	40.2	907	3	US-08-990-140-4	Sequence 4, Appli
43	33	40.2	940	4	US-08-810-712-7	Sequence 7, Appli
44	33	40.2	1299	4	US-08-460-900C-62	Sequence 62, Appli
45	33	40.2	1434	2	US-08-540-406-10	Sequence 10, Appli
46	33	40.2	1434	3	US-08-656-055-10	Sequence 10, Appli
47	33	40.2	1434	4	US-08-954-668-10	Sequence 10, Appli
48	33	40.2	1434	5	PCT-US95-13233-10	Sequence 10, Appli
49	33	40.2	1447	3	US-08-540-406-19	Sequence 19, Appli
50	33	40.2	1447	3	US-08-656-055-19	Sequence 19, Appli
51	33	40.2	1447	4	US-08-954-668-19	Sequence 19, Appli
52	33	40.2	1447	5	PCT-US95-13233-19	Sequence 19, Appli
53	33	40.2	2500	3	US-08-801-263A-2	Sequence 2, Appli
54	33	40.2	2500	3	US-09-102-248-2	Sequence 2, Appli
55	33	40.2	2512	2	US-08-801-263A-9	Sequence 9, Appli
56	33	40.2	2512	3	US-09-102-248-9	Sequence 9, Appli
57	33	40.2	2517	2	US-08-801-263A-5	Sequence 5, Appli
58	33	40.2	2517	3	US-09-102-248-5	Sequence 5, Appli
59	32	39.0	20	2	US-08-564-972-21	Sequence 21, Appli
60	32	39.0	193	2	US-08-564-972-8	Sequence 8, Appli
61	32	39.0	184	1	US-08-148-058A-27	Sequence 27, Appli
62	32	39.0	194	1	US-08-148-058A-29	Sequence 29, Appli
63	32	39.0	194	1	US-08-478-042-27	Sequence 27, Appli
64	32	39.0	194	1	US-08-478-042-29	Sequence 29, Appli
65	32	39.0	194	2	US-08-645-215-27	Sequence 27, Appli

ALIGNMENTS

RESULT 1
US-08-346-849-15
; Sequence 15, Application US/08346849
; Patent No. 5670483
; GENERAL INFORMATION:
; APPLICANT: Zhang, Shuquang
; APPLICANT: Lockshin, Curtis
; APPLICANT: Rich, Alexander
; APPLICANT: Holmes, Todd
; TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
; TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,849
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/973,326
FILING DATE: 28 DECEMBER 1992
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: MIT-6008
TELEPHONE: (617) 861-9540
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-346-849-15

Query Match 48.8%; Score 40; DB 1; Length 51;
Best Local Similarity 56.2%; Pred. No. 1.5;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 YVPPLLLEGVGEKFM 16
DB 17 YHPDLNLEPGAEELFL 32

RESULT 2
US-08-293-284A-15
Sequence 15, Application US/08293284A
Patent No. 5955343
GENERAL INFORMATION:
APPLICANT: Holmes, Todd
APPLICANT: Zhang, Shuguang
APPLICANT: Rippstein, Michael
APPLICANT: Lockshin, Curtis
TITLE OF INVENTION: STABLE MACROSCOPIC MEMBRANES FORMED BY
TITLE OF INVENTION: SELF-ASSEMBLY OF AMPHIPHILIC PEPTIDES AND USES
NUMBER OF SEQUENCES: 64
SEQUENCE CHARACTERISTICS:
ADDRESS: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA: 08/293,284A
FILING DATE: 22 AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/973,326
ATTORNEY/AGENT INFORMATION:
FILING DATE: 28-DEC-1992
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: MIT-6008A
TELEPHONE: (617) 861-9540
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 51 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-293-284A-15

Query Match 48.8%; Score 40; DB 2; Length 51;
Best Local Similarity 56.2%; Pred. No. 1.5;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 YVPPLLLEGVGEKFM 16
DB 17 YHPDLNLEPGAEELFL 32

RESULT 3
US-09-413-814-46
Sequence 46, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bock, Peter
APPLICANT: Bock, Peter
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberger, Steven L
APPLICANT: Hofle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Mueller, Joachim
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
TITLE OF INVENTION: heteropolyketide compounds
FILE REFERENCE: PCT/US 95/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
SEQUENCE CHARACTERISTICS:
LENGTH: 1544
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-46

Query Match 47.6%; Score 39; DB 4; Length 1544;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 PPLLEVG 10
DB 816 PPLMLEVG 823

RESULT 4
US-08-381-936-2
Sequence 2, Application US/08381936
Patent No. 5792923
GENERAL INFORMATION:
APPLICANT: ROBER, Manuela
APPLICANT: GEIDER, Gebhardt
APPLICANT: GEIDER, Klaus
APPLICANT: GEIDER, Klaus
TITLE OF INVENTION: DNA sequences which lead to the
TITLE OF INVENTION: formation of polyfructans (levans), plasmids containing
TITLE OF INVENTION: these sequences as well as a process for preparing
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESS: Sorangium cellulosum, Faber, Gorb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US

ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/381,936
FILING DATE: 09-FEB-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 93 02110
FILING DATE: 09-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P42270618
FILING DATE: 08-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meliman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-108
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-381-936-2

Query Match 43.9%; Score 36; DB 1; Length 415;
Best Local Similarity 46.2%; Pred. No. 94;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
QY 2 VPPLLEVGVEEK 14
Db 273 LPPLITAVGVNDQ 285

RESULT 5
US-08-943-374-2
Sequence 2, Application US/08943374
Patent No. 6028249
GENERAL INFORMATION:
APPLICANT: ROBER, Manuela
APPLICANT: GEIER, Gebhardt
APPLICANT: GEIDER, Klaus
APPLICANT: WILLMITZER, Lothar
TITLE OF INVENTION: DNA sequences which lead to the
formation of polyfructans (levans), plasmids containing
these sequences as well as a process for preparing
transgenic plants.
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,374
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/381,936
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P42270618
FILING DATE: 08-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meliman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-108
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-943-374-2

Query Match 43.9%; Score 36; DB 3; Length 415;
Best Local Similarity 46.2%; Pred. No. 94;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEK 14
Db 273 LPPLITAVGVNDQ 285

RESULT 6
US-08-749-902-7
Sequence 7, Application US/08749902
Patent No. 5985635
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goli, Surya K.
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: NOVEL HUMAN SERINE/THREONINE
PROTEIN KINASES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/749,902
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0150 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 189508
US-08-749-902-7

Query Match 43.3%; Score 35.5; DB 2; Length 525;
Best Local Similarity 60.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 YVPPLLLEGVGEK 15
DB 413 YVAPSVLE-SVKERF 426

RESULT 7
US-08-749-902-8
Sequence 8, Application US/08749902
Patent No. 5985635

GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya Kifer L.
TITLE OF INVENTION: NOVEL HUMAN SERINE/THREONINE
TITLE OF INVENTION: PROTEIN KINASES
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 1174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/749,902
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0150 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-4165
TELEFAX: 415-845-4165

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULAR TYPE: protein
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1562
US-08-749-902-8

Query Match 43.3%; Score 35.5; DB 2; Length 525;
Best Local Similarity 60.0%; Pred. No. 1.5e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 YVPPLLLEGVGEK 15
DB 413 YVAPSVLE-SVKERF 426

RESULT 8

US-08-938-548B-7
Sequence 7, Application US/08938548B
Patent No. 6001358

GENERAL INFORMATION:
APPLICANT: Yanagisawa, Masashi
APPLICANT: Bergsma, Derk
APPLICANT: Wilson, Shelagh
APPLICANT: Brooks, David
TITLE OF INVENTION: NOVEL LIGANDS OF THE NEUROPEPTIDE
TITLE OF INVENTION: RECEPTOR HFGAN72
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: United States of America
ZIP: 19406

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,548B
FILING DATE: 26-SEPT-1997
CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/887,382
FILING DATE: 2-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,519
FILING DATE: 19-MAR-1997
APPLICATION NUMBER: 60/033,604
FILING DATE: 18-DEC-1997
NAME: Elizabeth J. Hecht
REGISTRATION NUMBER: 41,824
REFERENCE/DOCKET NUMBER: ATG50037-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5009
TELEFAX: 610-270-5090

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULAR TYPE: protein
US-08-938-548B-7

Query Match 42.7%; Score 35; DB 3; Length 32;
Best Local Similarity 54.5%; Pred. No. 7;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLLEGVGE 12
DB 21 LPPALLSLGVD 31

RESULT 9
US-08-554-840-13
Sequence 13, Application US/08554840
Patent No. 6001358

GENERAL INFORMATION:
APPLICANT: BLACK, Amelia
APPLICANT: HANNA, Nadia
APPLICANT: HANNA, Ricardo A.
APPLICANT: NEWMAN, Roland A.
TITLE OF INVENTION: HUMANIZED ANTIBODIES TO HUMAN gp39,

;; TITLE OF INVENTION: COMPOSITIONS CONTAINING AND THERAPEUTIC USE THEREOF
;; NUMBER OF SEQUENCES: 28
;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Burns, Doane, Swecker & Mathis
;; STREET: P.O. Box 1404
;; CITY: Alexandria
;; STATE: Virginia
;; COUNTRY: United States
;; ZIP: 22313-1404

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patent In Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/554,840

;; FILING DATE: 07-NOV-1995

;; CLASSIFICATION: 536

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Teskin, Robin L.

;; REGISTRATION NUMBER: 35,030

;; REFERENCE/DOCKET NUMBER: 012712-127

;; TELEPHONE: (703) 836-6620

;; TELEFAX: (703) 836-2021

;; INFORMATION FOR SEQ ID NO: 13:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 80 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; US-08-554-840-13

Query Match 42.7%; Score 35; DB 3; Length 80;

Best Local Similarity 53.8%; Pred. No. 21;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 PPLLEVGVEERF 15

Db 32 PPKLLIYGVDRF 44

RESULT 10

US-08-938-548B-10

; Sequence 10, Application US/08938548B

; Patent No. 6001963

; GENERAL INFORMATION:

; APPLICANT: Yanagisawa, Masashi

; APPLICANT: Bergsma, Derk

; APPLICANT: Wilson, Shelagh

; APPLICANT: Brooks, David

; APPLICANT: Gellai, Miklos

; TITLE OF INVENTION: NOVEL LIGANDS OF THE NEUROPEPTIDE

; TITLE OF INVENTION: RECEPTOR HFGAN72

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

; STATE: PA

; COUNTRY: United States of America

; ZIP: 19406

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/938,548B

; FILING DATE: 26-SEPT-1997

; CLASSIFICATION: 536

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/887,382
;; FILING DATE: 2-JUL-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/820,519
;; FILING DATE: 19-MAR-1997
;; APPLICATION NUMBER: 60/033,604
;; FILING DATE: 17-DEC-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Elizabeth J. Hecht
;; REGISTRATION NUMBER: 41,824
;; REFERENCE/DOCKET NUMBER: ATG50037-2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 610-270-5009
;; TELEFAX: 610-270-5090
;; TELEX:

;; INFORMATION FOR SEQ ID NO: 10:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 123 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: protein

;; US-08-938-548B-10

Query Match 42.7%; Score 35; DB 3; Length 123;

Best Local Similarity 54.5%; Pred. No. 34;

Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 VPPLLEVGVE 12

Db 14 LPPALLSLGVD 24

RESULT 11

US-08-938-548B-6

; Sequence 6, Application US/08938548B

; Patent No. 6001963

; GENERAL INFORMATION:

; APPLICANT: Yanagisawa, Masashi

; APPLICANT: Bergsma, Derk

; APPLICANT: Wilson, Shelagh

; APPLICANT: Brooks, David

; APPLICANT: Gellai, Miklos

; TITLE OF INVENTION: NOVEL LIGANDS OF THE NEUROPEPTIDE

; TITLE OF INVENTION: RECEPTOR HFGAN72

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SmithKline Beecham Corporation

; STREET: 709 Swedeland Road

; CITY: King of Prussia

; STATE: PA

; COUNTRY: United States of America

; ZIP: 19406

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/938,548B

; FILING DATE: 26-SEPT-1997

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/887,382

; FILING DATE: 2-JUL-1997

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/820,519

; FILING DATE: 19-MAR-1997

; APPLICATION NUMBER: 60/033,604

; FILING DATE: 17-DEC-1997

; ATTORNEY/AGENT INFORMATION:

NAME: Elizabeth J. Hecht
REGISTRATION NUMBER: 61,824
REFERENCE/DOCKET NUMBER: ATG50037-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5009
TELEFAX: 610-270-5090
TELEX: 610-270-5090
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 130 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-938-548B-6

Query Match 42.7%; Score 35; DB 3; Length 130;
Best Local Similarity 54.5%; Pred. NO. 36;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 VPPLLEVGVE 12
DB 21 LPPALLSLGVD 31

RESULT 12 895-26
US-08-876-885-26 Application US/08876885
Sequence 26 6174713
Patent No. 6174713

GENERAL INFORMATION:
APPLICANT: Shen, Xiaoyu
APPLICANT: Houman, Fariba
TITLE OF INVENTION: CANDIDA CYTOPLASMIC TRYPTOPHANYL-tRNA
TITLE OF INVENTION: SAME
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PC-Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,885
FILING DATE: 16-JUN-1997

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mary J.
REGISTRATION NUMBER: E. 22,592
REFERENCE/DOCKET NUMBER: CPI97-02

TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:
LENGTH: 130 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

Query Match 42.7%; Score 35; DB 4; Length 424;
Best Local Similarity 46.2%; Pred. NO. 15e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 4 PLLLEVCVERKFM 16
DB 129 PLVIELTDDKFL 141

RESULT 13

US-08-432-016-6
Sequence 6, Application US/08432016
Patent No. 5998172
GENERAL INFORMATION:
APPLICANT: HAYNES, BARTON F.
APPLICANT: ARUFFO, ALEJANDRO
APPLICANT: PATEL, DHAVALKUMAR
APPLICANT: BOWEN, MICHAEL A.
APPLICANT: MARQUARDT, HANS
APPLICANT: WILSON, MARY J.
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHUYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22201-4714

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PC-Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,016
FILING DATE: 01-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA: US 08/333,350
APPLICATION NUMBER: 08/333,350
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA: US 08/143,903
APPLICATION NUMBER: 08/143,903
FILING DATE: 02-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: E. 22,592
REFERENCE/DOCKET NUMBER: 1579-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:
LENGTH: 130 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-432-016-6

Query Match 42.7%; Score 35; DB 2; Length 486;
Best Local Similarity 72.7%; Pred. NO. 1.7e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 VPPLLEVGVE 12
DB 448 VPELLETGVE 458

RESULT 14

US-08-684-594-6
Sequence 6, Application US/08684594
Patent No. 5998172
GENERAL INFORMATION:
APPLICANT: HAYNES, BARTON F.
APPLICANT: ARUFFO, ALEJANDRO
APPLICANT: PATEL, DHAVALKUMAR

APPLICANT: BOWEN, MICHAEL A.
APPLICANT: MARQUARDT, HANS
TITLE OF INVENTION: CD6 LIGAND
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/684,594
FILING DATE: 18-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/432,016
FILING DATE: 01-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/333,350
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/143,903
FILING DATE: 02-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 1579-112
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: linear
MOLECULE TYPE: protein
US-08-684-594-6

Query Match 42.7%; Score 35; DB 2; Length 486;
Best Local Similarity 72.7%; Pred. No. 1.7e+02;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVE 12
Db 448 VTPELETGVE 458

RESULT 15
PCT-US94-00198-3
Sequence 3, Application PC/TUS9400198
GENERAL INFORMATION:
APPLICANT: Schering Corp.
TITLE OF INVENTION: RAS Associated GAP Proteins
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schering Corp.
STREET: 1 Giraldo Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 6.0.8

SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00198
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,824
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lunn, Paul G.
REGISTRATION NUMBER: 32,743
REFERENCE/DOCKET NUMBER: DX0352 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)822-7255
TELEFAX: (201)822-7039
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2938 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
PCT-US94-00198-3

Query Match 42.7%; Score 35; DB 5; Length 2938;
Best Local Similarity 58.3%; Pred. No. 1.4e+03;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLEVGVEEKFM 16
Db 2762 LLEVGFEDELV 2773

RESULT 16
US-09-042-012-15
Sequence 15, Application US/09042012A
Patent No. 6111087
GENERAL INFORMATION:
APPLICANT: RETHWILM, Axel
APPLICANT: LINDEMANN, Dirk
APPLICANT: WINTER, Jan
TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
FILE REFERENCE: 032751-006
CURRENT APPLICATION NUMBER: US/09/042.012A
CURRENT FILING DATE: 1998-03-13
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 55
TYPE: PRT
ORGANISM: transmembrane anchor domain of HFV env protein
US-09-042-012-15

Query Match 41.5%; Score 34; DB 4; Length 55;
Best Local Similarity 54.5%; Pred. No. 20;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 YVPLLEVGVE 11
Db 24 YLKPILIGVG 34

RESULT 17
US-09-042-012-9
Sequence 9, Application US/09042012A
Patent No. 6111087
GENERAL INFORMATION:
APPLICANT: RETHWILM, Axel
APPLICANT: LINDEMANN, Dirk

```

; APPLICANT: WINTER, Jan
; TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
; FILE REFERENCE: 032751-006
; CURRENT APPLICATION NUMBER: US/09/042,012A
; CURRENT FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 63
; TYPE: PRT
; ORGANISM: transmembrane anchor domain of HFV env protein
US-09-042-012-9

```

```

Query Match      41.5%; Score 34; DB 4; Length 63;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 YVPPLLLEGV 11
   1 1 1 1 1 1 1
Db 32 YLKPILIGV 42

```

```

RESULT 18
US-09-042-012-11
; Sequence 11, Application US/09042012A
; Patent No. 6111087
; GENERAL INFORMATION:
; APPLICANT: RETHWILM, Axel
; APPLICANT: LINDEMANN, Dirk
; APPLICANT: WINTER, Jan
; TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
; FILE REFERENCE: 032751-006
; CURRENT APPLICATION NUMBER: US/09/042,012A
; CURRENT FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 63
; TYPE: PRT
; ORGANISM: fusion of the transmembrane anchor domains of the HFV and HIV
US-09-042-012-11

```

```

Query Match      41.5%; Score 34; DB 4; Length 63;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 YVPPLLLEGV 11
   1 1 1 1 1 1 1
Db 32 YLKPILIGV 42

```

```

RESULT 19
US-09-042-012-7
; Sequence 7, Application US/09042012A
; Patent No. 6111087
; GENERAL INFORMATION:
; APPLICANT: RETHWILM, Axel
; APPLICANT: LINDEMANN, Dirk
; APPLICANT: WINTER, Jan
; TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
; FILE REFERENCE: 032751-006
; CURRENT APPLICATION NUMBER: US/09/042,012A
; CURRENT FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 64
; TYPE: PRT
; ORGANISM: wildtype HFV proteins
US-09-042-012-7

```

```

Query Match      41.5%; Score 34; DB 4; Length 64;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 YVPPLLLEGV 11
   1 1 1 1 1 1 1
Db 33 YLKPILIGV 43

```

```

RESULT 20
US-09-305-086-1
; Sequence 1, Application US/09305086
; Patent No. 6111087
; GENERAL INFORMATION:
; APPLICANT: LINDEMANN, Dirk
; APPLICANT: RETHWILM, Axel
; TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
; FILE REFERENCE: 032751-010
; CURRENT APPLICATION NUMBER: US/09/305,086
; CURRENT FILING DATE: 1998-05-04
; EARLIER APPLICATION NUMBER: US 08/816,439
; EARLIER FILING DATE: 1997-03-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 64
; TYPE: PRT
; ORGANISM: wildtype HFV proteins
US-09-305-086-1

```

```

Query Match      41.5%; Score 34; DB 4; Length 64;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 YVPPLLLEGV 11
   1 1 1 1 1 1 1
Db 33 YLKPILIGV 43

```

```

RESULT 21
US-09-042-012-17
; Sequence 17, Application US/09042012A
; Patent No. 6111087
; GENERAL INFORMATION:
; APPLICANT: RETHWILM, Axel
; APPLICANT: LINDEMANN, Dirk
; APPLICANT: WINTER, Jan
; TITLE OF INVENTION: EXPRESSION OF A FOAMY VIRUS ENVELOPE PROTEIN
; FILE REFERENCE: 032751-006
; CURRENT APPLICATION NUMBER: US/09/042,012A
; CURRENT FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 71
; TYPE: PRT
; ORGANISM: fusion of cytoplasmic tail of HFV env protein with transmembrane
US-09-042-012-17

```

```

Query Match      41.5%; Score 34; DB 4; Length 71;
Best Local Similarity 54.5%; Pred. No. 27;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 YVPPLLLEGV 11
   1 1 1 1 1 1 1
Db 24 YLKPILIGV 34

```

```

RESULT 22
US-08-946-026-1

```

; Sequence 1, Application US/08946026
; Patent No. 6034218
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Mitcham, Jennifer L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/946,026
; FILING DATE: 07-OCT-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.424C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 89 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-946-026-1

Query Match 41.5%; Score 34; DB 3; Length 89;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 LLEVGVEERF 15
I: :|:|:|:|
Db 26 LMSIGLDERF 35

RESULT 23
US-08-081-929-10
; Sequence 10, Application US/08081929
; Patent No. 6160090
; GENERAL INFORMATION:
; APPLICANT: Schllessinger, Joseph
; APPLICANT: Barnea, Gilad
; APPLICANT: Grumet, Martin H.
; APPLICANT: Margolis, Richard U.
; TITLE OF INVENTION: A NEW CLASS OF RPTases: THEIR
; TITLE OF INVENTION: STRUCTURAL DOMAINS AND LIGANDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/081,929
; FILING DATE: 23-JUN-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-041-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-081-929-10

Query Match 41.5%; Score 34; DB 4; Length 260;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 LLEVGVEE 13
I: :|:|:|:|
Db 151 ILFEIGVEE 159

RESULT 24
US-08-937-972-6
; Sequence 6, Application US/08937972
; Patent No. 5932443
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: ANTIGENS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/937,972
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0400 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:

LENGTH: 271 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 199586
US-08-937-972-6

Query Match 41.5%; Score 34; DB 2; Length 271;
Best Local Similarity 46.2%; Pred. No. 1.3e+02;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 VPPLLEVGVEEK 14
DB 237 VEPILSVGEDE 249

RESULT 25
US-08-129-8 Application US/08129129
; Sequence 3, 5932443
; Patent No. 5932443
; GENERAL INFORMATION:
; APPLICANT: DE SILVA, Jacqueline
; APPLICANT: SAFFORD, Richard
; APPLICANT: HUGHES, Stephen Glyn
; TITLE OF INVENTION: PLANT PROMOTER INVOLVED IN
; TITLE OF INVENTION: CONTROLLING LIPID BIOSYNTHESIS IN SEEDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA: Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/129,129
; FILING DATE: 18-JAN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91303098.7
; FILING DATE: 08-APR-1991
; PRIOR APPLICATION DATA: PCT/GB92/00627
; APPLICATION NUMBER: PCT/GB92/00627
; FILING DATE: 08-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, Paul N.
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 203424/T7016(C)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627CUSH
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-129-8

Query Match 41.5%; Score 34; DB 1; Length 315;
Best Local Similarity 37.5%; Pred. No. 1.6e+02;
Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 YVPPLLEVGVEEKFM 16
DB 228 FIPLLKADAGVVKRFI 243

RESULT 26
US-08-937-972-3 Application US/08937972
; Sequence 3, 5932443
; Patent No. 5932443
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: OXYGEN-INDUCED ANTIGENS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA: Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/937,972
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0400 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 321 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BLADNOT04
; CLONE: 1318190
US-08-937-972-3

Query Match 41.5%; Score 34; DB 2; Length 321;
Best Local Similarity 46.2%; Pred. No. 1.6e+02;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 VPPLLEVGVEEK 14
DB 237 VEPILSVGEDE 249

RESULT 27
US-08-993-260-3 Application US/08993260
; Sequence 3, 6031089
; Patent No. 6031089
; GENERAL INFORMATION:
; APPLICANT: Gajewski, Michael J.
; TITLE OF INVENTION: NO. 6031089el Sequences of p56, a Proteins Which
; TITLE OF INVENTION: Affects K-ATP Channels
; NUMBER OF SEQUENCES: 4

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pharmacia and Upjohn, Co., Intel. Prop. Law
;; ADDRESS: (1920-32-LAW)
;; STREET: 301 Henrietta Street
;; CITY: Kalamazoo
;; STATE: Michigan
;; COUNTRY: U.S.A.
;; ZIP: 49001
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/993,260
;; FILING DATE:
;;
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wootton, Thomas A.
;; REGISTRATION NUMBER: 35,004
;; REFERENCE/DOCKET NUMBER: 6092
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (616) 833-7914
;; TELEFAX: (616) 833-8897
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 494 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: N-terminal
;;
US-08-993-260-3

Query Match 41.5%; Score 34; DB 3; Length 494;
Best Local Similarity 54.5%; Pred. No. 2.7e+02;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 LLEVGVVEKFM 16
II:III::I:
DB 195 LLOVGTQRFI 205

RESULT 28
US-08-946-026-3
; Sequence 3, Application US/08946026
; Patent No. 6034218
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Mitcham, Jennifer L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,026
; FILING DATE: 07-OCT-1997

;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Maki, David J.
;; REGISTRATION NUMBER: 31,392
;; REFERENCE/DOCKET NUMBER: 210121.424C1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (206) 622-4900
;; TELEFAX: (206) 682-6031
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 858 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;;
US-08-946-026-3

Query Match 41.5%; Score 34; DB 3; Length 858;
Best Local Similarity 50.0%; Pred. No. 5.1e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 6 LLEVGVVEKFP 15
II:III::I:
DB 253 LMSIGLDEKF 262

RESULT 29
US-08-540-406-4
; Sequence 4, Application US/08540406
; Patent No. 5837538
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHEW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: JOHNSON, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/540,406
; FILING DATE: 06-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: a60190-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-540-406-4

Query Match 41.5%; Score 34; DB 2; Length 1311;
Best Local Similarity 46.7%; Pred. No. 8.4e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEKFM 16
DB 496 VPFLALGLGVQDMFL 510

RESULT 30
US-08-656-055-4
; Sequence 4, Application US/08656055
; Patent No. 6027882
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHW P
; APPLICANT: SCOTT, CH, LISA V
; APPLICANT: GOODRICH, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; STATE: San Francisco
; COUNTRY: CA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; SOFTWARE: PC DOS/MS-DOS
; CURRENT APPLICATION DATA:
; FILING DATE: 08/540.406
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: a60190-1
; TELEPHONE: 415-781-1989
; TELEFAX: 415-781-1989
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-656-055-4

Query Match 41.5%; Score 34; DB 3; Length 1311;
Best Local Similarity 46.7%; Pred. No. 8.4e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEKFM 16
DB 496 VPFLALGLGVQDMFL 510

RESULT 31
US-08-954-668-4
; Sequence 4, Application US/08954568
; Patent No. 6172200
; GENERAL INFORMATION:
; APPLICANT: SCOTT, MATHW P
; APPLICANT: GOODRICH, LISA V
; APPLICANT: GOODRICH, RONALD L
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley, Hoag & Eliot
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; SOFTWARE: PC DOS/MS-DOS
; CURRENT APPLICATION DATA:
; FILING DATE: 20-Oct-1997
; CLASSIFICATION INFORMATION:
; AT: US/08954568
; NAME: Vincent, Matthew P
; REGISTRATION NUMBER: 36709
; REFERENCE/DOCKET NUMBER: SUV-003.06
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-954-668-4

Query Match 41.5%; Score 34; DB 4; Length 1311;
Best Local Similarity 46.7%; Pred. No. 8.4e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEKFM 16
DB 496 VPFLALGLGVQDMFL 510

RESULT 32
US-08-13233-4
; Sequence 4, Application PC/TUS9513233
; Patent No. 6172200
; GENERAL INFORMATION:
; APPLICANT: THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY
; TITLE OF INVENTION: Patched Genes and their Use
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; SOFTWARE: PC DOS/MS-DOS
; CURRENT APPLICATION DATA:
; FILING DATE: 06-OCT-1990
; CLASSIFICATION INFORMATION:
; AT: US/08954568
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: a60190-1
; TELEPHONE: 415-781-1989
; TELEFAX: 415-781-1989
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1311 amino acids
; TYPE: amino acid
US-08-13233-4

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-13233-4

Query Match 41.5%; Score 34; DB 5; Length 1311;
Best Local Similarity 46.7%; Pred. No. 8.4e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLLEVGVEEFK 16
DB 496 VPFLALGLGVQDMFL 510

RESULT 33

US-09-413-304-15

; Sequence 15, Application US/09413304

; Patent No. 6207387

; GENERAL INFORMATION:

; APPLICANT: Louis J. Elsas II

; APPLICANT: K. Muralidharan

; TITLE OF INVENTION: MOLECULAR DIAGNOSTICS FOR GALACTOSEMIA

; FILE REFERENCE: 05010.0079

; CURRENT APPLICATION NUMBER: US/09/413,304

; CURRENT FILING DATE: 1999-10-06

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 15

; LENGTH: 379

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:/No. 6207387e =

US-09-413-304-15

Query Match

40.9%; Score 33.5; DB 4; Length 379;

Best Local Similarity 56.2%; Pred. No. 2.4e+02;

Matches 9; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1 YVPPLLLEVGVEEFK 16

DB 322 YVPPLLRSATV-RKFM 336

RESULT 34

US-08-905-223-320

; Sequence 320, Application US/08905223

; Patent No. 6222029

; GENERAL INFORMATION:

; APPLICANT: Edwards, Jean-Baptiste D.

; APPLICANT: Duclert, Aymeric

; APPLICANT: Lacroix, Bruno

; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS

; NUMBER OF SEQUENCES: 503

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe, Martens, Olson & Bear

; STREET: 501 West Broadway

; CITY: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92101-3505

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: Win95

; SOFTWARE: Word

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/905,223

; FILING DATE:

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 320:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 125 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: LINEAR

; MOLECULE TYPE: PROTEIN

; ORIGINAL SOURCE:

; ORGANISM: Homo Sapiens

; TISSUE TYPE: Brain

; FEATURE:

; NAME/KEY: sig_peptide

; LOCATION: -18..-1

; IDENTIFICATION METHOD: Von Heijne matrix

; OTHER INFORMATION: score 5.6

; OTHER INFORMATION: seq WHFLASFFPRAGC/HG

US-08-905-223-320

Query Match

40.2%; Score 33; DB 4; Length 125;

Best Local Similarity 53.8%; Pred. No. 80;

Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 PPLLEVGVEEFK 15

DB 59 PTILLVVGPAEQF 71

RESULT 35

US-08-977-767-1

; Sequence 1, Application US/08977767

; Patent No. 5972684

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Yue, Henry

; APPLICANT: Greenwald, Sara

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: CARBONIC ANHYDRASE VIII

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/977,767

; FILING DATE: Herewith

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0423 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

LENGTH: 328 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: OVARNOT03
 CLONE: 2059155
 US-08-977-767-1

Query Match 40.2%; Score 33; DB 2; Length 328;
 Best Local Similarity 46.2%; Pred. No. 2.5e+02;
 Matches 6; Conservative 3; Mismatches 4; Indels 0;

OY 1 YVPLLEVGVEE 13
 DB 83 FLPLRLSTGGER 95

RESULT 36
 US-08-974-546-5
 : Sequence 8; Application US/08974546
 : Patent No. 5945287
 : GENERAL INFORMATION:
 : APPLICANT: Hillman, Jennifer L.
 : APPLICANT: Lal, Preeti
 : APPLICANT: Corley, Neil C.
 : TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
 : NUMBER OF SEQUENCES: 6
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Incyte Pharmaceuticals, Inc.
 : STREET: 3174 Porter Dr.
 : CITY: Palo Alto
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94304
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/974,546
 : FILING DATE: Filed Herewith
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Billings, Lucy J.
 : REGISTRATION NUMBER: 36,749
 : REFERENCE/DOCKET NUMBER: PF-0428
 : TELEPHONE: 650-853-0555
 : TELEFAX: 650-853-0555
 : INFORMATION FOR SEQ ID NO: 5:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 340 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : IMMEDIATE SOURCE:
 : LIBRARY: G6B10K
 : CLONE: 1816452
 US-08-974-546-5

Query Match 40.2%; Score 33; DB 2; Length 340;
 Best Local Similarity 53.3%; Pred. No. 2.6e+02;
 Matches 8; Conservative 0; Mismatches 7; Indels 0;

OY 1 YVPLLEVGVEEKF 15
 DB 83 FLPLRLSTGGER 95

DB 31 YHDKNKEPGAEEKF 45

RESULT 37
 US-08-987-904A-2
 : Sequence 2; Application US/08987904A
 : Patent No. 6027917
 : GENERAL INFORMATION:
 : APPLICANT: Celeste, Anthony J.
 : APPLICANT: Murray, Beth
 : TITLE OF INVENTION: BONE MORPHOGENETIC PROTEIN (BMP)- 17 AND BMP-18
 : NUMBER OF SEQUENCES: 7
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Genetics Institute, Inc.
 : STREET: 87 CambridgePark Drive
 : CITY: Cambridge
 : STATE: MA
 : COUNTRY: US
 : ZIP: 02140
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/987,904A
 : FILING DATE: 10/03/97
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: LAZAR, STEVEN R
 : REGISTRATION NUMBER: 32,618
 : REFERENCE/DOCKET NUMBER: GI 5307
 : TELEPHONE: (617) 876-8581
 : TELEFAX: (617) 876-8581
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 366 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : IMMEDIATE SOURCE:
 : MEDIUM TYPE: Floppy disk
 US-08-987-904A-2

Query Match 40.2%; Score 33; DB 3; Length 366;
 Best Local Similarity 46.2%; Pred. No. 2.8e+02;
 Matches 6; Conservative 4; Mismatches 3; Indels 0;

OY 4 PLLEVGVEEKF 16
 DB 196 PLLQVSVQREHL 208

RESULT 38
 US-08-987-904A-4
 : Sequence 4; Application US/08987904A
 : Patent No. 6027917
 : GENERAL INFORMATION:
 : APPLICANT: Celeste, Anthony J.
 : APPLICANT: Murray, Beth
 : TITLE OF INVENTION: BONE MORPHOGENETIC PROTEIN (BMP)- 17 AND BMP-18
 : NUMBER OF SEQUENCES: 7
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Genetics Institute, Inc.
 : STREET: 87 CambridgePark Drive
 : CITY: Cambridge
 : STATE: MA
 : COUNTRY: US
 : ZIP: 02140
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/987,904A
FILING DATE: 10-DEC-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, STEVEN R
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: GI 5307
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8769
TELEFAX: (617) 876-8581
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-987-904A-4

Query Match 40.2%; Score 33; DB 3; Length 366;
Best Local Similarity 46.2%; Pred. No. 2.8e+02;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEKFM 16

DB 196 PLLLVSVQREHL 208

RESULT 39

US-09-080-044-5
Sequence 5, Application US/090800044
Patent No. 6074649
GENERAL INFORMATION:
APPLICANT: AUDONNET, Jean-Christophe F.
APPLICANT: BAUDU, Philippe G.
APPLICANT: RIVIERE, Michel A.
TITLE OF INVENTION: RECOMBINANT VACCINE CONTAINING FELINE HERPES VIRUS TYPE
TITLE OF INVENTION: 1, PARTICULARLY FOR TREATING FELINE INFECTIOUS
FILE REFERENCE: AUDONNET
CURRENT APPLICATION NUMBER: US/09/080,044
CURRENT FILING DATE: 1998-05-15
EARLIER APPLICATION NUMBER: PCT/FR96/01830
EARLIER FILING DATE: 1996-11-19
EARLIER APPLICATION NUMBER: 95/14450
EARLIER FILING DATE: 1995-11-30
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 395
TYPE: PRT
ORGANISM: Feline herpesvirus 1
US-09-080-044-5

Query Match 40.2%; Score 33; DB 3; Length 395;
Best Local Similarity 54.5%; Pred. No. 3.1e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLLEVGVEE 13

DB 4 PPSRLEVGIN 14

RESULT 40

US-09-080-983-5
Sequence 5, Application US/09080983
Patent No. 6197948
GENERAL INFORMATION:

APPLICANT: Zhu, Hai-Ying
APPLICANT: Ling, Kai-Shu
APPLICANT: Gonsalves, Dennis
TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS TYPE 2 PROTEINS
TITLE OF INVENTION: AND THEIR USES
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,983
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/047,194
FILING DATE: 20-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1631
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 459 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-080-983-5

Query Match 40.2%; Score 33; DB 4; Length 459;
Best Local Similarity 37.5%; Pred. No. 3.7e+02;
Matches 6; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEKFM 16

DB 299 YADDICTDMGFETKFM 314

RESULT 41

US-08-609-230A-9
Sequence 9, Application US/08609230A
Patent No. 5866333
GENERAL INFORMATION:
APPLICANT: Innerarity, Thomas L.
APPLICANT: Qian, Xiaobing
APPLICANT: Yamanaka, Shinya
TITLE OF INVENTION: Screening Methods to Detect mRNA Targets
TITLE OF INVENTION: of Editing Enzymes
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT FILING DATE: US/08/609,230A
;; APPLICATION NUMBER: US/08/609,230A
;; FILING DATE: 01-MAR-1995
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fitts, Renee A.
;; REGISTRATION NUMBER: 35,136
;; TELEPHONE: 650-326-2400
;; TELEFAX: 650-326-2422
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 906 amino acids
;; TYPE: amino acids
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-609-230A-9

Query Match 40.2%; Score 33; DB 2; Length 906;
Best Local Similarity 63.6%; Pred No. 8.2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 LLEVGVEEKF 16
DB 100 LLNVGVESKLI 110

RESULT 42

US-08-990-140-4

;; Sequence 4, Application US/08990140A
;; Patent No. 609,935
;; GENERAL INFORMATION:
;; APPLICANT: Olsen, Henrik S.
;; APPLICANT: Ruben, Steven M.
;; APPLICANT: Sonenberg, Nahum
;; APPLICANT: Methot, Nathalie
;; APPLICANT: Rom, Eran
;; TITLE OF INVENTION: Human prt1-like Subunit Protein (hPrtl) and Human
;; FILE REFERENCE: 1488-070001
;; CURRENT FILING DATE: 1997-12-12
;; EARLIER APPLICATION NUMBER: US/08/990,140A
;; EARLIER FILING DATE: 1996-12-13
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 4
;; LENGTH: 907
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; US-08-990-140-4

Query Match 40.2%; Score 33; DB 3; Length 907;
Best Local Similarity 63.6%; Pred. NO. 8.2e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 LLEVGVEEKF 16
DB 100 LLNVGVESKLI 110

RESULT 43

US-08-810-712-7

;; Sequence 7, Application US/08810712G
;; Patent No. 616,0106
;; GENERAL INFORMATION:
;; APPLICANT: Yeda Research and Development Co. LTD
;; TITLE OF INVENTION: Tumor Suppressor Genes, Proteins Encoded Thereby and
;; TITLE OF INVENTION: Use of said Genes and Proteins

;; FILE REFERENCE: sequencelist
;; CURRENT FILING DATE: US/08/810,712G
;; APPLICATION NUMBER: US/08/810,712G
;; EARLIER APPLICATION NUMBER: PCT/US94/11598
;; EARLIER FILING DATE: 1994-10-12
;; NUMBER OF SEQ ID NOS: 31
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 7
;; LENGTH: 940
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; US-08-810-712-7

Query Match 40.2%; Score 33; DB 4; Length 940;
Best Local Similarity 63.6%; Pred. NO. 8.1e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 LLEVGVEEKF 16
DB 133 LLNVGVESKLI 143

RESULT 44

US-08-460-900C-62
;; Sequence 62, Application US/08460900C
;; Patent No. 6165747
;; GENERAL INFORMATION:
;; APPLICANT: Tadmor, Phillip W.
;; APPLICANT: McMahon, Andrew P.
;; APPLICANT: Tabin, Clifford J.
;; APPLICANT: Bumcrot, David A.
;; APPLICANT: Marti-Gorostiza, Elisa
;; TITLE OF INVENTION: Vertebbrate Embryonic Pattern-Inducing
;; FILE REFERENCE: PCT/US94/11598
;; NUMBER OF SEQUENCES: 62
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
;; STREET: One Post Office Square
;; CITY: Boston
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02109

COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SEQUENCE CHARACTERISTICS:
CURRENT APPLICATION DATA: US/08/460,900C
FILING DATE: 5-JUNE-1995
PRIOR APPLICATION DATA: US 08/435,093
FILING DATE: 14-DEC-1994
APPLICATION NUMBER: US 08/356,060
PRIOR APPLICATION DATA: US 08/176,427
FILING DATE: 30-DEC-1993
APPLICATION NUMBER: 36,709
NAME/VISCENT Matthew P.
REGISTRATION NUMBER: HMV-006.05
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
SEQUENCE CHARACTERISTICS:
LENGTH: 1299 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

Search completed: June 28, 2001, 11:56:09
Job time: 190 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 42.81 Seconds
(without alignments)
28.470 Million cell updates/sec

Title: US-09-439-313-554
Perfect score: 82
Sequence: 1 YVPPLLLEVGVEEKFM 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	54.9	299	2 PT0060	N-acetylphosphinotriptide-deacetylase
2	45	54.9	299	2 A47031	Streptomyces viridochromogenes
3	43	52.4	545	2 J00341	Streptomyces viridochromogenes
4	42	51.2	367	2 B72394	intercellular adhe
5	42	51.2	449	2 T17419	citrate synthase -
6	42	51.2	528	2 T31905	probable alkylhal
7	42	51.2	700	2 A83434	hypothetical prote
8	41	50.0	237	2 G5084	conserved hypothet
9	41	50.0	260	1 J50635	hypothetical prote
10	41	50.0	309	2 G83383	rRNA (adenine-N6-)
11	41	50.0	436	2 B70473	probable esterase/
12	41	50.0	3472	2 T31308	protoporphyrinogen
13	40	48.8	108	2 S66990	hypothetical 367K
14	40	48.8	171	2 E86820	probable membrane
15	40	48.8	243	2 A69449	16S rRNA processin
16	40	48.8	339	2 S20062	heme biosynthesis
17	40	48.8	348	2 B72493	heat shock protein
18	40	48.8	687	2 G81970	hypothetical prote
19	40	48.8	687	2 E81027	probable glycine--
20	40	48.8	737	2 C70132	glycyl-tRNA synthet
21	40	48.8	916	2 D83093	hypothetical prote
22	40	48.8	2114	2 E96505	secretion protein
23	39	47.6	162	2 B84152	hypothetical prote
24	39	47.6	297	2 S49885	hypothetical prote
25	39	47.6	333	2 T48739	probable membrane
26	39	47.6	419	2 T29201	probable atp-speci
27	39	47.6	441	2 F86708	hypothetical prote
28	39	47.6	531	2 T23835	citrate (sl)-synth
29	39	47.6	570	2 E84825	hypothetical prote
					probable protein k

30	39	47.6	606	2 B81729	glutamine--fructos
31	39	47.6	722	2 A82617	glycyl-tRNA synthet
32	39	47.6	800	2 T26683	hypothetical prote
33	39	47.6	840	2 G85648	probable usher pro
34	39	47.6	1963	2 T49914	callose synthase c
35	38	46.3	227	2 C83046	probable transcrip
36	38	46.3	277	2 T34993	probable oxidoredu
37	38	46.3	387	2 G72288	conserved hypothet
38	38	46.3	419	2 T15088	hypothetical prote
39	38	46.3	449	2 S76839	hypothetical prote
40	38	46.3	490	2 T26171	hypothetical prote
41	38	46.3	506	2 S32561	cysteine proteinas
42	38	46.3	561	2 F75191	hypothetical prote
43	38	46.3	635	2 B72215	hypothetical prote
44	38	46.3	700	2 F64078	translation elonga
45	38	46.3	941	2 T37626	DNA mismatch repai
46	38	46.3	1930	2 F86200	protein Fl2K11.17
47	38	46.3	2140	2 T18543	probable cell-adhe
48	38	46.3	2561	1 I40456	peptide synthetase
49	37	45.1	115	1 VKLJND	trans-regulatory s
50	37	45.1	139	2 B64010	hypothetical prote
51	37	45.1	164	2 D84152	hypothetical prote
52	37	45.1	179	2 E65134	hypothetical prote
53	37	45.1	179	2 B86004	hypothetical prote
54	37	45.1	207	2 C36961	hypothetical prote
55	37	45.1	225	2 E69350	hypothetical prote
56	37	45.1	266	2 T36341	probable esterase
57	37	45.1	288	2 G84782	hypothetical prote
58	37	45.1	297	2 G81329	probable curved-DN
59	37	45.1	433	1 ITHUC	alpha-1-antichymot
60	37	45.1	437	2 T30897	hypothetical prote
61	37	45.1	446	2 T35627	probable integral
62	37	45.1	450	2 S01367	inner membrane pro
63	37	45.1	559	2 T33168	hypothetical prote
64	37	45.1	596	2 T23193	hypothetical prote
65	37	45.1	621	2 A84933	glutamine--fructos

ALIGNMENTS

RESULT 1	
PT0060	N-acetylphosphinotriptide-deacetylase - Streptomyces viridochromogenes
C:Species: Streptomyces viridochromogenes	
C:Date: 31-Mar-1992 #sequence_revision 26-Apr-1996 #text_change 19-May-2000	
C:Accession: S20686; PT0060	
R:Alijah, R.; Hillemann, D.; Nussbaumer, B.; Pelzer, S.; Wohleben, W.	
submitted to the EMBL Data Library, March 1992	
A:Description: Gene disruption and gene replacement analysis of a 4 kb BamHI fragment	
A:Reference number: S20683	
A:Accession: S20686	
A>Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-299 <ALI>	
A:Cross-references: EMBL:X65195; NID:g47997; PIDN:CAA46315.1; PID:g48001	
R:Wohleben, W.; Arnold, W.; Broer, I.; Hillemann, D.; Strauch, E.; Puehler, A.	
Gene 70, 25-37, 1988	
A:Title: Nucleotide sequence of the phosphinothricin N-acetyltransferase gene from St	
A:Reference number: J0409; MUID:89196914	
A:Accession: PT0060	
A:Molecule type: DNA	
A:Residues: 1-164 <NOH>	
A:Cross-references: GB:M22827; NID:g295177; PIDN:AAA72710.1; PID:g295180	
A:Experimental source: strain Tue 494	
C:Superfamily: probable lipolytic protein ybaC	

Query Match 54.9%; Score 45; DB 2; Length 299;
Best Local Similarity 66.7%; Pred. No. 2.6;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 VPPLLLEVGVEE 13

Db 226 LPPLLIQVGSSE 237

RESULT 2

A:Title: Nucleotide sequence analysis reveals linked N-acetyl hydrolase, thioesterase, and
 A:Reference number: A47031; MUID:91294191
 A:Accession: A47031
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-299 <RAI>
 A:Note: Sequence extracted from NCBI backbone (NCBIN:41300, NCBI:41303)
 C:Superfamily: probable lipolytic protein ybac

Query Match 54.9%; Score 45; DB 2; Length 299;
 Best Local Similarity 66.7%; Pred. No. 2.6;
 Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 VPPLLEVGVEE 13
 Db 226 LPPLLIQVGSSE 237

RESULT 3

A:Title: Cellular adhesion molecule 1 - rat
 A:Reference number: J00341
 A:Accession: S21765; J00341
 A:Molecule type: mRNA
 A:Status: preliminary
 A:Residues: 1-545 <KIT>
 A:Note: Sequence extracted from NCBI backbone (NCBIN:9220778, PIDN:BAA00759.1; PID:9220779)
 C:Superfamily: intercellular adhesion molecule; immunoglobulin homology

Query Match 52.4%; Score 43; DB 2; Length 545;
 Best Local Similarity 61.5%; Pred. No. 1.1;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 4 PPLLEVGVEE 16
 Db 224 PPLLEVGVEE 236

RESULT 4

A:Title: citrate synthase - Thermotoga maritima (strain MSB8)
 A:Reference number: B72394
 A:Accession: B72394
 A:Molecule type: protein
 A:Status: preliminary
 A:Residues: 1-323-329, 1999
 A:Note: Sequence extracted from NCBI backbone (NCBIN:41300, NCBI:41303)
 C:Superfamily: citrate synthase

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
 A:Reference number: A72200; MUID:99287316
 A:Accession: B72394
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-367 <ARN>
 A:Note: Sequence extracted from NCBI backbone (NCBIN:41300, NCBI:41303)
 C:Superfamily: citrate (si)-synthase

Query Match 51.2%; Score 42; DB 2; Length 367;
 Best Local Similarity 53.8%; Pred. No. 1.1;
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 2 VPPLLEVGVEE 14
 Db 218 VPPLLEIGSEDR 230

RESULT 5

A:Title: Probable alkylhalidase (EC 3.8.1.1) plta - Pseudomonas fluorescens
 A:Reference number: T17419
 A:Accession: T17419
 A:Molecule type: protein
 A:Status: preliminary
 A:Residues: 1-449 <NOW>
 A:Note: Sequence extracted from NCBI backbone (NCBIN:41300, NCBI:41303)
 C:Superfamily: alkylhalidase

Query Match 51.2%; Score 42; DB 2; Length 449;
 Best Local Similarity 72.7%; Pred. No. 1.4;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 4 PPLLEVGVEE 14
 Db 54 PPLLEIGVMEK 64

RESULT 6

A:Title: Ribosomal protein T05H4.13 - Caenorhabditis elegans
 A:Reference number: T31905
 A:Accession: T31905
 A:Molecule type: protein
 A:Status: preliminary
 A:Residues: 1-528 <BLA>
 A:Note: Sequence extracted from NCBI backbone (NCBIN:41300, NCBI:41303)
 C:Superfamily: ribosomal protein

Query Match 52.4%; Score 43; DB 2; Length 545;
 Best Local Similarity 61.5%; Pred. No. 1.1;
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 4 PPLLEVGVEE 16
 Db 224 PPLLEVGVEE 236

Query Match 51.2%; Score 42; DB 2; Length 528;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPLLEVGVEKFM 16
: : : : :
Db 316 YIPPTVLDVEKSDPFM 331

RESULT 7
A:Species: Streptomyces lividans
C:Species: Streptomyces lividans
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 05-May-2000
A:Accession: JS0635
R:Jenkins, G.; Cundliffe, E.
Gene 108, 55-62, 1991
A:Title: Cloning and characterization of two genes from Streptomyces lividans that co
A:Reference number: JS0635; MUID:92104506
A:Accession: JS0635
A:Molecule type: DNA
A:Residues: 1-260 <JEN>
A:Cross-references: GB:M74717; NID:g153345; PIDN:AAA26779.1; PID:g153346
A:Experimental source: strain TK21
C:Comment: This enzyme confers resistance to lincomycin.
C:Comment: This enzyme catalyzes the monomethylation of a specific adenosine within 2
C:Genetics:
A:Gene: lrm
C:Superfamily: rRNA (adenine-N6-)-methyltransferase
C:Keywords: methyltransferase; S-adenosylmethionine

Query Match 50.0%; Score 41; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PPLLEVG 10
: : : : :
Db 36 PPLLEVG 43

RESULT 10
G83383
Probable esterase/deacetylase PA2098 [imported] - Pseudomonas aeruginosa (strain PA01
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
A:Accession: G83383
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: G83383
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-309 <STO>
A:Cross-references: GB:AE004637; GB:AE004091; NID:g9948105; PIDN:AAG05486.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2098
C:Superfamily: probable lipolytic protein ybaC

Query Match 50.0%; Score 41; DB 2; Length 309;
Best Local Similarity 58.3%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEE 13
: : : : :
Db 232 LPPLLVQVGEDE 243

RESULT 11
B70473
protoporphyrinogen oxidase - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
A:Accession: B70473
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666

Query Match 51.2%; Score 42; DB 2; Length 700;
Best Local Similarity 57.1%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEERF 15
: : : : :
Db 559 VPULLIAPGIQERF 572

RESULT 8
G65084
hypothetical protein b2985 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
A:Accession: G65084
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G65084
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-237 <BLAT>
A:Cross-references: GB:AE000381; GB:U00096; NID:g2367181; PIDN:AAC76021.1; PID:g1789358;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: conserved hypothetical protein b2986

Query Match 50.0%; Score 41; DB 2; Length 237;
Best Local Similarity 42.9%; Pred. No. 10;
Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 YVPLLEVGVEERK 14
: : : : :
Db 168 YLPVLLIRLGIDEQ 181

RESULT 9
JS0635

Query Match	48.8%	Score 40;	DB 2;	Length 108;
Best Local Similarity	40.0%	Pred. NO.	6.5;	
Matches	6;	Conservative	6;	Mismatches
				3;
				Indels
				0;
				Gaps
				0;

A:Reference number: S20062; MUID:92093635
 A:Accession: S20062
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-339 <RAAL>
 A:Cross-references: EMBL:X62421
 R:Raabe, T.; Manley, J.L.
 Submitted to the EMBL Data Library, October 1991
 A:Description: Primary structure of a human homologue to the Escherichia coli DnaJ protein
 A:Reference number: S18086
 A:Accession: S18086
 A:Molecule type: mRNA
 A:Residues: 1-2, 'K', 4-10, 'Q', 12-33, 'K', 35, 'K', 37-42, 'K', 44, 'K', 46-58, 'K', 60-243, 'N', 245-
 A:Cross-references: EMBL:X62421; NID:g30850; PIDN:CAA44287.1; PID:g30851
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:4-67/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 48.8%; Score 40; DB 2; Length 339;
 Best Local Similarity 56.2%; Pred. No. 23;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEEKF 16
 I I I I I I I I I
 DB 30 YHPDLNLEPGAEELFL 45

RESULT 17
 B72493
 hypothetical protein APE2585 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: B72493
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
 A:Reference number: A72450; MUID:99310339
 A:Accession: B72493
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-348 <KAW>
 A:Cross-references: DBJ:AP000064; NID:g5105945; PIDN:BAA81602.1; PID:d1045388; PID:g510
 A:Experimental source: strain K1.
 C:Genetics:
 A:Gene: APE2585

Query Match 48.8%; Score 40; DB 2; Length 348;
 Best Local Similarity 81.8%; Pred. No. 24;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLLEVGVEEKF 15
 I I I I I I I I I
 DB 98 LLLRVGVVEYF 108

RESULT 18
 G81970
 probable glycine-tRNA ligase (EC 6.1.1.14) beta chain NMA0523 [imported] - Neisseria me
 C:Species: Neisseria meningitidis
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: G81970
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
 i; Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A:Reference number: A81775; MUID:20222556
 A:Accession: G81970
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-687 <PAR>
 A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83815.1; PID:g737920

A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: glyS; NMA0523
 C:Superfamily: glycine-tRNA ligase beta chain
 C:Keywords: ligase

Query Match 48.8%; Score 40; DB 2; Length 687;
 Best Local Similarity 53.8%; Pred. No. 52;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEKF 16
 I I I I I I I I I
 DB 259 PVVLEAGFEEHFL 271

RESULT 19
 E81027
 glycyl-tRNA synthetase, beta chain NMB1930 [imported] - Neisseria meningitidis (strai
 C:Species: Neisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: E81027
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755
 A:Accession: E81027
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-687 <TET>
 A:Cross-references: GB:AE002541; GB:AE002098; NID:g727175; PIDN:AAF42259.1; PID:g722
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:
 A:Gene: NMB1930
 C:Superfamily: glycine-tRNA ligase beta chain

Query Match 48.8%; Score 40; DB 2; Length 687;
 Best Local Similarity 53.8%; Pred. No. 52;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEKF 16
 I I I I I I I I I
 DB 259 PVVLEAGFEEHFL 271

RESULT 20
 C70132
 hypothetical protein BB0259 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C:Accession: C70132
 R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Yu
 ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.
 A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943
 A:Accession: C70132
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-737 <LE>
 A:Cross-references: GB:AE001136; GB:AE000783; NID:g2688152; PIDN:AAC66641.1; PID:g268
 A:Experimental source: strain B31

Query Match 48.8%; Score 40; DB 2; Length 737;
 Best Local Similarity 50.0%; Pred. No. 56;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

RESULT 23
 h; Biological protein BH4018 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: B84152
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000 the alkaliphilic bacterium Bacillus halodurans a
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans
 A:Accession: B84152
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-162 <STO>
 A:Cross-references: GB:BA001520; GB:BA000004; NID:910176401; PIDN:BA807737.1; GSPDB:G
 A:Experimental source: strain C-125
 A:Genetic: 0; Conservative 0; Indels 0; Gaps 0;
 A:Gene: BH4018

 Query Match 47.6% Score 39; DB 2; Length 162;
 Best Local Similarity 54.3%; Pred. No. 16;
 Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

 QY 4 PLLEVGVEEK 14
 DB 151 PVLIEGLEQK 161
 |:|:|:|:|
 |:|:|:|:|

 RESULT 24
 S49885
 probable membrane protein YIL124w - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein Y18277.05.
 C:Species: Saccharomyces cerevisiae
 C:Date: 13-Jan-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
 C:Accession: S49885
 R:Hamlyn, N.; Churcher, C.
 submitted to the EMBL Data Library, November 1994
 A:Reference number: S49881
 A:Accession: S49885
 A:Molecule type: DNA
 A:Residues: 1-297 <HAM>
 A:Cross-references: GB:Z47047; EMBL:Z46833; NID:G603997; PID:G763222; GSPDB:GN000009;
 C:Genetic: 0; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 A:Map position: 9L
 A:Gene: MIPS:YIL124w
 C:Superfamily: retinol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Keywords: transmembrane protein.
 F:131/Domain: short-chain alcohol dehydrogenase homology <SADH>
 F:255-271/Domain: transmembrane status predicted <TM>

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-333 <SCH>
 A:Cross-references: EMBL:AL353819; GSPDB:GN00112; NCSP:8D4.130
 A:Experimental source: cosmid contig 8D4; strain 74
 C:Genetics:
 A:Gene: NCSP:8D4.130
 A:Map position: 2
 A:Introns: 56/2; 66/1; 108/2; 138/1
 C:Superfamily: succinate-CoA ligase (ADP-forming) alpha chain

Query Match 47.6%; Score 39; DB 2; Length 333;
 Best Local Similarity 53.8%; Pred. No. 35;
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLLEGVVEE 13
 ||| | |
 Db 106 FVPPPLAAGIEE 118

RESULT 26
 T29201
 hypothetical protein T03F1.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29201
 R:Du, Z.; Le, T.T.
 submitted to the EMBL Data Library, February 1997
 A:Description: The sequence of C. elegans cosmid T03F1.
 A:Reference number: Z20586
 A:Accession: T29201

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-419 <DUZ>
 A:Cross-references: EMBL:U88169; PIDN:AAB42231.1; GSPDB:GN00019; CESP:T03F1.1
 A:Experimental source: strain Bristol N2; clone T03F1
 C:Genetics:
 A:Gene: CESP:T03F1.1
 A:Map position: 1
 A:Introns: 147/2; 238/3

Query Match 47.6%; Score 39; DB 2; Length 419;
 Best Local Similarity 46.2%; Pred. No. 45;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLLEGVVEEK 14
 ||||: |:
 Db 240 VPPLVVASGIDER 252

RESULT 27
 F86708
 citrate (si)-synthase (EC 4.1.3.7) [imported] - Lactococcus lactis subsp. lactis (strain
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 04-Apr-2001
 C:Accession: F86708
 R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrlich
 Genome Res. in press, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium.
 A:Reference number: AB6625
 A:Accession: F86708
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-441 <STO>
 A:Cross-references: GB:AE005176; NID:gl2723577; PIDN:AAK04768.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: gltA
 C:Superfamily: citrate (si)-synthase
 C:Keywords: carbon-carbon lyase; coenzyme A; oxo-acid-lyase

Query Match 47.6%; Score 39; DB 2; Length 441;
 Best Local Similarity 53.3%; Pred. No. 48;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLLEGVVEEK 15
 || |:
 Db 335 YVETLVKRGLEEEF 349

RESULT 28
 T23835
 hypothetical protein M88.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T23835
 R:Sulston, J.
 submitted to the EMBL Data Library, June 1994
 A:Reference number: Z19806
 A:Accession: T23835
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-531 <WIL>
 A:Cross-references: EMBL:Z34802; PIDN:CAA84336.1; GSPDB:GN00021; CESP:M88.1
 A:Experimental source: clone M88
 C:Genetics:
 A:Gene: CESP:M88.1
 A:Map position: 3
 A:Introns: 30/3; 71/1; 220/2; 251/3; 288/3; 381/3; 437/2
 C:Superfamily: glucuronosyltransferase

Query Match 47.6%; Score 39; DB 2; Length 531;
 Best Local Similarity 53.3%; Pred. No. 59;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLLEGVVEEK 15
 |||||: |:
 Db 187 YVPLMWESDDEMGF 201

RESULT 29
 E84825
 Probable protein kinase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: E84825
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: E84825
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-570 <STO>
 A:Cross-references: GB:AE002093; NID:g6598931; PIDN:AAF18726.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g40120
 A:Map position: 2

Query Match 47.6%; Score 39; DB 2; Length 570;
 Best Local Similarity 54.5%; Pred. No. 64;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 PLLLEGVVEEK 14
 | | |:
 Db 188 PFLDIGLEDK 198

RESULT 30

C:Accession: T49914
R:Bevan, M.; Robben, J.; Grymonprez, B.; Volckaert, G.; Bancroft, I.; Mewes, H.W.; Rudd,
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25024
A:Accession: T49914
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1963 <BEY>
A:Cross-references: EMBL:AL353013; GSPDB:GN00063; ATSP:T24H18.170
A:Experimental source: cultivar Columbia; BAC clone T24H18
C:Genetics:
A:Gene: ATSP:T24H18.170
A:Map position: 5
A:Introns: 66/1; 99/3; 141/2; 172/3; 209/3; 248/3; 279/3; 322/2; 364/3; 410/3; 436/2; 46
1110/2; 1214/1; 1237/1; 1300/2; 1350/3; 1390/3; 1430/2; 1458/2; 1496/1; 1572/3; 1806/1;
Query Match 47.6%; Score 39; DB 2; Length 1963;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 4 PILLLEGVVEKF 15
Db 1628 PMLMEIGLGRGF 1639
RESULT 35
C83046
Probable transcription regulator PA4806 [imported] - Pseudomonas aeruginosa (strain PA01
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83046
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: C83046
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-227 <STO>
A:Cross-references: GB:AE004893; GB:AE004091; NID:g9951063; PIDN:AAG08192.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4806
Query Match 46.3%; Score 38; DB 2; Length 227;
Best Local Similarity 46.7%; Pred. No. 34;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 YVPPLLLEGVVEKF 15
Db 9 HAPLLYERGIARF 23
RESULT 36
T34993
Probable oxidoreductase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T34993
R:Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z21564
A:Accession: T34993
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-277 <SEE>
A:Cross-references: EMBL:AL034443; PIDN:CAA22355.1; GSPDB:GN00070; SCOEDB:SC4B5.01c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC4B5.01c
C:Superfamily: aldehyde reductase
Query Match 46.3%; Score 38; DB 2; Length 277;
Best Local Similarity 63.6%; Pred. No. 43;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 2 VPPLLELVGVE 12
Db 5 VPPIILNGVE 15
RESULT 37
G72288
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: G72288
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: G72288
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-387 <ARN>
A:Cross-references: GB:AE001772; GB:AE000512; NID:g4981693; PIDN:AAD36223.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TW1147
C:Superfamily: sensory transduction system regulatory protein; response regulator hom
Query Match 46.3%; Score 38; DB 2; Length 387;
Best Local Similarity 50.0%; Pred. No. 63;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 1 YVPPLLLEGVGVE 12
Db 14 WIRPLLTIQIGVE 25
RESULT 38
T15088
hypothetical protein K06A5.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
C:Accession: T15088
R:Wamsley, P.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid K06A5.
A:Reference number: Z18291
A:Accession: T15088
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-419 <WAM>
A:Cross-references: EMBL:AF039038; NID:g2736359; PID:g2736364; PIDN:AAB94172.1; GSPDB
A:Experimental source: strain Bristol N2; clone K06A5
C:Genetics:
A:Gene: CESP:K06A5.6
A:Map position: 1
A:Introns: 84/3; 265/1; 313/3
C:Superfamily: acyl-CoA dehydrogenase
Query Match 46.3%; Score 38; DB 2; Length 419;
Best Local Similarity 46.2%; Pred. No. 68;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 2 VPPLLELVGVEEK 14

DB 129 VAPLIQLGTED 141

RESULT 39

S76839
 A:Title: A hypothetical protein - Synechocystis sp. (strain PCC 6803)
 A:Reference number: S76839
 A:Accession: S76839
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-449 <K>
 A:Cross-references: GB:AB001339; NID:gl653836; PIDN:BAAL8751.1; PID:gl65384
 C:Superfamily: arsenical pump membrane protein

Query Match 46.3%; Score 38; DB 2; Length 449;
 Best Local Similarity 54.5%; Pred. No. 74;
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VEPILLEVGE 12
 DB 126 IPELAQEGVD 136

RESULT 40

S76839
 A:Title: A hypothetical protein W04G5.4 - Caenorhabditis elegans
 A:Reference number: W04G5.4
 A:Accession: W04G5.4
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-490 <WIL>
 A:Cross-references: EMBL:293391; PIDN:CAB07681.1; GSPDB:GN00019; CESP:W04G5.4
 C:Superfamily: arsenical pump membrane protein

Query Match 46.3%; Score 38; DB 2; Length 490;
 Best Local Similarity 37.5%; Pred. No. 81;
 Matches 6; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEKPM 16
 DB 48 YLKPIELPNIENVL 63

RESULT 41

S32561
 A:Title: A hypothetical protein - Plasmodium vinckei
 A:Reference number: S32561
 A:Accession: S32561
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-506 <ROS>
 A:Cross-references: GB:AE001814; GB:AE000512; NID:g4982332; PIDN:AAD36826.1; PID:g498

Query Match 46.3%; Score 38; DB 2; Length 506;
 Best Local Similarity 43.8%; Pred. No. 84;
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEKPM 16
 DB 387 YGPTVAVGASEDFV 402

RESULT 42

F75191
 A:Title: A hypothetical protein PAB0027 - Pyrococcus abyssi (strain Orsay)
 A:Reference number: PAB0027
 A:Accession: P75191
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-561 <KAM>
 A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB48981.1; PID:g545

Query Match 46.3%; Score 38; DB 2; Length 561;
 Best Local Similarity 43.8%; Pred. No. 95;
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEKPM 16
 DB 45 YVAKRLSEIGIEYFL 60

RESULT 43

B72215
 A:Title: A hypothetical protein TM1762 - Thermotoga maritima (strain MSB8)
 A:Reference number: B72215
 A:Accession: B72215
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-635 <ARN>
 A:Cross-references: GB:AE001814; GB:AE000512; NID:g4982332; PIDN:AAD36826.1; PID:g498

Query Match 46.3%; Score 38; DB 2; Length 635;
 Best Local Similarity 72.7%; Pred. No. 11e+02;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 PPLLEVGVEE 13
| | | | | | | |
Db 366 PERLLEVGVEE 376

Search completed: June 28, 2001, 11:56:57
Job time: 238 sec

RESULT 44
F64078
translation elongation factor EF-G - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 02-Feb-2001
C:Accession: F64078
R:Flaeschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630
A:Accession: F64078
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-700 <TIGR>
A:Cross-references: GB:U32739; GB:L42023; NID:gl573559; PIDN:AAC2237.1; PID:gl573567; T
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog
C:Keywords: GTP binding; nucleotide binding; P-loop; protein biosynthesis
F:11-145/Domain: translation elongation factor Tu homology <ETU>
F:17-24/Region: nucleotide-binding motif A (P-loop)
F:142-145/Region: GTP-binding NKXD motif
F:250-252/Region: GTP-binding SAK/L motif
F:23,24,62,142,143,145,250/Binding site: Mg-GTP (Lys, Thr, Asn, Lys, Asp, Ser) #sta

Query Match 46.3%; Score 38; DB 2; Length 700;
Best Local Similarity 66.7%; Pred. NO. 1.2e+02;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 PPLLEVGVEEXF 15
| | | | | | | |
Db 168 PLQLPVGAENF 179

RESULT 45
T37626
DNA mismatch repair protein homolog - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T38256; T37626
R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A:Reference number: Z21781
A:Accession: T38256
A:Molecule type: DNA
A:Residues: 168-941 <BRO>
A:Cross-references: EMBL:Z98559; PIDN:CAB11169.1; GSPDB:GN00066; SPDB:SPAC23C11.18C
R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21733
A:Accession: T37626
A:Molecule type: DNA
A:Residues: 1-209 <BR2>
A:Cross-references: EMBL:Z99091; PIDN:CAB11774.1; GSPDB:GN00066; SPDB:SPAC13F5.01C
A:Experimental source: strain 972h-; cosmid c13F5
C:Genetics:
A:Gene: SPAC23C11.18C; SPDB:SPAC13F5.01C
A:Map position: 1

Query Match 46.3%; Score 38; DB 2; Length 941;
Best Local Similarity 53.8%; Pred. NO. 1.7e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 22.61 Seconds
(without alignments)
24.241 Million cell updates/sec

Title: US-09-439-313-554
Perfect score: 82
Sequence: 1 YVPPLLLEVGVEEKFM 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	54.9	299	1 BAH_STRHY	Q01109 streptomyces
2	43	52.4	545	1 ICAL1 RAT	Q00238 rattus norv
3	41	50.0	237	1 YGHS_ECOLI	Q46843 escherichia
4	39	47.6	297	1 YIM4_YEAST	P40471 saccharomyc
5	39	47.6	722	1 SYGB_XYLFA	Q9pc26 xytelia fas
6	38	46.3	449	1 Y753_SYNY3	P74635 synchocyst
7	38	46.3	506	1 CYSP_PLAVN	P46102 plasmodium
8	38	46.3	699	1 EFG_HAEIN	P43925 haemophilus
9	38	46.3	941	1 MSH1_SCHPO	O13921 schizosacch
10	38	46.3	2561	1 PPS1_BACSU	P39845 bacillus su
11	37	45.1	116	1 REV_HVIND	P18803 human immun
12	37	45.1	139	1 Y589_HAEIN	P44020 haemophilus
13	37	45.1	179	1 YRFC_ECOLI	P45752 escherichia
14	37	45.1	207	1 YACE_PSEPU	P36644 pseudomonas
15	37	45.1	423	1 RACT_HUMAN	P01011 homo sapien
16	37	45.1	450	1 ENVZ_SALTY	P08982 salmonella
17	37	45.1	451	1 PPOX_MYCLE	Q50008 mycobacteri
18	37	45.1	608	1 GLMS_BACAI	P57138 b glucosani
19	37	45.1	679	1 SYGB_BACSU	P54381 bacillus su
20	37	45.1	1491	1 AT7A_MOUSE	Q64430 mus musculu
21	37	45.1	1492	1 A77A_RAT	P70705 rattus norv
22	37	45.1	2405	1 D7HA_CHLRE	Q39610 chlamydomon
23	36.5	44.5	483	1 Y045_MYCGE	P47291 mycoplasma
24	36	43.9	144	1 Y850_HAEIN	P44060 haemophilus
25	36	43.9	248	1 Y1PA_YEAST	P53039 saccharomyc
26	36	43.9	262	1 FHUF_ECOLI	P39405 escherichia
27	36	43.9	267	1 PIV6_ADE40	P48309 human adeno
28	36	43.9	306	1 KHSE_SYNY3	P73646 synchocyst
29	36	43.9	365	1 MAP3_SCHPO	P31397 schizosacch
30	36	43.9	368	1 DNAJ_BACSH	Q69269 bacillus sp
31	36	43.9	373	1 CATB_RHOOP	P95608 rhodococcus
32	36	43.9	373	1 EGON_DRONE	P15370 drosophila
33	36	43.9	377	1 DNAJ_LISMO	Q955a3 listeria mo

34	36	43.9	379	1 DNAJ_LACLA	P35514 lactococcus
35	36	43.9	393	1 YSH8_CAEEL	Q09949 caenorhabdi
36	36	43.9	415	1 SACB_ERWAM	Q46654 erwinia amy
37	36	43.9	547	1 SPAG_HUMAN	Q9uew8 homo sapien
38	36	43.9	553	1 SPAG_RAT	Q88506 rattus norv
39	36	43.9	556	1 SPAG_MOUSE	Q921w9 mus musculu
40	36	43.9	580	1 IUCC_ECOLI	Q47318 escherichia
41	36	43.9	608	1 GLMS_CHLPN	Q926u0 c glucosami
42	36	43.9	703	1 EFG_ECOLI	P02996 escherichia
43	36	43.9	732	1 ACPH_HUMAN	P13798 homo sapien
44	36	43.9	930	1 DPO1_HAEIN	P43741 haemophilus
45	36	43.9	985	1 INVA_YERPS	P11922 yersinia ps
46	36	43.9	1220	1 PTC1_BRARE	Q98864 brachydanio
47	36	43.9	1226	1 METH_ECOLI	P13009 escherichia
48	36	43.9	3587	1 SRF2_BACSU	Q04747 bacillus su
49	35.5	43.3	502	1 K6B1_HUMAN	P23443 homo sapien
50	35.5	43.3	502	1 K6B1_RAT	P21425 rattus norv
51	35.5	43.3	843	1 MVPA_DICDI	P34118 dictyosteli
52	35.5	43.3	3083	1 POLG_ZYMVR	Q89330 z genome po
53	35	42.7	72	1 RL15_BACLI	P35138 bacilli li
54	35	42.7	100	1 URE3_MYCTU	P50043 mycobacteri
55	35	42.7	130	1 OREX_MOUSE	O55241 mus musculu
56	35	42.7	130	1 OREX_RAT	O55232 rattus norv
57	35	42.7	140	1 FUCU_ECOLI	P11555 escherichia
58	35	42.7	146	1 RL15_BACST	P04452 bacillus st
59	35	42.7	146	1 RL15_BACSU	P19946 bacillus su
60	35	42.7	173	1 DCD_ACIAM	Q02103 acidianus a
61	35	42.7	194	1 CYCY_BRAJA	P30960 bradyrhizob
62	35	42.7	201	1 YD07_NPVAC	P24650 autographa
63	35	42.7	209	1 YDGI_BACSU	P96707 bacillus su
64	35	42.7	253	1 TPIS_BORBU	Q59182 borrelia bu
65	35	42.7	254	1 TPIS_CHLPN	Q926j6 chlamydia p

ALIGNMENTS

RESULT 1	BAH_STRHY	STANDARD;	PRT;	299 AA.
ID	BAH_STRHY	STANDARD;	PRT;	299 AA.
AC	Q01109;			
DC	01-APR-1993 (Rel. 25, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DE	01-FEB-1996 (Rel. 33, Last annotation update)			
GN	ACETYL-HYDROLASE (EC 3.1.1.-)			
OS	Streptomyces hygroscopicus.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;			
OC	Actinomycetales; Streptomycetaceae; Streptomyces.			
OX	NCBI_TaxID=1912;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-ATCC 21705;			
RX	MEDLINE-91294191; PubMed-2066341;			
RA	Raibaud A., Zalacain M., Holt T.G., Tizard R., Thompson C.J.;			
RT	"Nucleotide sequence analysis reveals linked N-acetyl hydrolase,			
RT	thioesterase, transport, and regulatory genes encoded by the			
RT	Bialaphos biosynthetic gene cluster of Streptomyces hygroscopicus."			
RL	J. Bacteriol. 173:4454-4463(1991).			
CC	-!- FUNCTION: THIS PROTEIN REMOVES THE N-ACETYL GROUP FROM			
CC	BIALAPHOS AS ONE OF THE FINAL STEPS OF THE BIALAPHOS			
CC	BIOSYNTHETIC PATHWAY.			
CC	-!- PATHWAY: BIALAPHOS BIOSYNTHESIS.			
CC	-!- SIMILARITY: BELONGS TO THE "GDHG" FAMILY OF LIPOLYTIC ENZYMES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			

EMBL: M64783; AAN79277.1; -
 InterPro: IPR002168; -
 PROSITE: PS01173; LIPASE_GDXG_HIS; 1.
 PROSITE: PS01174; LIPASE_GDXG_SER; 1.
 Hydrolase; Antibiobiotic biosynthesis.
 FT ACT_SITE 73 73 POTENTIAL.
 FT ACT_SITE 143 143 POTENTIAL.
 SQ SEQUENCE 259 AA; 32096 MW; 4265C8E6E10FAE97 CRC64;
 Query Match 54.9%; Score 45; DB 1; Length 299;
 Best Local Similarity 66.7%; Pred. No. 1.3; Mismatches 1; Indels 0; Gaps 0;
 Matches 8; Conservative 3;
 QY 2 VPALLEVGVEE 13
 DB 226 LPPLLIVGSEE 237
 RESULT 2
 LOCAL RAT STANDARD; PRT; 545 AA.
 AC Q00238;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE INTERCELLULAR ADHESION MOLECULE-1 PRECURSOR.
 GN ICAM1 OR ICAM1-LIKE (Rat).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 RN [1]
 RP NCBI_TaxID=10116;
 SEQUENCE FROM N.A. PubMed-1349828;
 KIDUY Takashi T., Iigo Y., Yamatani T., Miyasaka M.,
 "Sequence and expression of rat ICAM-1.";
 RL Biochim. Biophys. Acta 1131:108-110(1992).
 CC -!- FUNCTION: ICAM PROTEINS ARE LIGANDS FOR THE LEUKOCYTE ADHESION
 MOLECULE-1 (CD11b) AND CD11c.
 CC -!- LEA-1 PROTEIN. CAUTION: TYPE I MEMBRANE PROTEIN.
 CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE ICAM FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use, but this statement is to be removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D00913; BAA00759.1; -
 DR PIR: S21765; S21765;
 DR RSP: J00341;
 DR InterPro: IPR003006; -
 DR Pfam: PF00047; 1g; 3.
 KW Immunoglobulin domain; Cell adhesion; Glycoprotein; Transmembrane;
 Repeat; Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 545
 FT DOMAIN 29 545
 FT TRANSMEM 493 517
 FT DOMAIN 518 545
 FT DOMAIN 41 103
 FT DOMAIN 128 193
 FT DOMAIN 230 297
 FT DOMAIN 325 389
 FT DOMAIN 428 495
 FT DISULFID 448 495
 FT DISULFID 52 96
 BY SIMILARITY.
 INTERCELLULAR ADHESION MOLECULE-1.
 POTENTIAL.
 CYTOPLASMIC (POTENTIAL).
 IG-LIKE C2-TYPE DOMAIN.
 IG-LIKE C2-TYPE DOMAIN.
 IG-LIKE C2-TYPE DOMAIN.
 IG-LIKE C2-TYPE DOMAIN.
 BY SIMILARITY.

FT DISULFID 135 186 BY SIMILARITY.
 FT DISULFID 237 290 BY SIMILARITY.
 FT DISULFID 332 382 BY SIMILARITY.
 FT DISULFID 430 469 BY SIMILARITY.
 FT SITE 177 179 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 47 183 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 183 193 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 202 202 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 309 309 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 344 344 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 396 396 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 464 464 N-LINKED (GLCNAC) (POTENTIAL).
 SQ SEQUENCE 545 AA; 60141 MW; 30F4546FAAD0CEFF4 CRC64;
 Query Match 52.4%; Score 43; DB 1; Length 545;
 Best Local Similarity 61.5%; Pred. No. 5.5; Mismatches 3; Indels 0; Gaps 0;
 Matches 8; Conservative 3;
 QY 4 PLLEVGVEEKF 16
 DB 224 PDLEVGTOQKFL 236
 RESULT 3
 YGHS_ECOLI STANDARD; PRT; 237 AA.
 ID YGHS_ECOLI
 AC Q46843;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 26.3 KDA ATP-BINDING PROTEIN IN GLCC-PITB INTERGENIC
 REGION.
 GN YGHS.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 RN [1]
 RP NCBI_TaxID=562;
 SEQUENCE FROM N.A. PubMed-9278503;
 STRAIN=K12 / MG1655;
 RX MEDLINE-97426617;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose B.J.,
 "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474(1997).
 RL
 CC -!- SIMILARITY: TO E.COLI YGHR AND YGHT.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use, but this statement is to be removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U28377; AAA69152.1;
 DR EMBL: AEO00000; AAC768.1;
 DR EMBL: G003003; G003003;
 KW Hypothetical protein; ATP-binding.
 FT NP_BIND 21 28
 SQ SEQUENCE 237 AA; 26346 MW; 69D8AE6673D7DA6F CRC64;
 Query Match 50.0%; Score 41; DB 1; Length 237;
 Best Local Similarity 42.9%; Pred. No. 5.2; Mismatches 6; Indels 0; Gaps 0;
 Matches 6; Conservative 6;

NCBI_TaxID=2371;
[1]
SEQUENCE FROM N.A.
STRAIN=9A5C;
MEDLINE=20365717; PubMed=10910347;
Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
Garnier M., Goldman G.H., Goldmann M.H.S., Gomes S.L., Gruber A.,
Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
Krieger J.E., Kutamae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
de Oliveira M.C. de Oliveira R.C., Palmieri D.A., Paris A.,
Peixoto R.B., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
Quaggio R.R., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
da Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
"The genome sequence of the plant pathogen *Xylella fastidiosa*";
Nature 406:151-159(2000).
-!- CATALYTIC ACTIVITY: ATP + L-GLYCINE + TRNA(GLY) = AMP +
PYROPHOSPHATE + L-GLYCYL-TRNA(GLY).
-!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS
(BY SIMILARITY).
-!- SUBCELLULAR LOCATION: CYTOPLASMIC.
-!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

EMBL; AF004015; AAF84761.1; -
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ SEQUENCE 722 AA; 80149 MW; AB6D0B4712B889EE CRC64;

Query Match 47.6%; Score 39; DB 1; Length 722;
Best Local Similarity 70.0%; Pred. No. 38;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PLLELVGVEE 13
|||:|:|
DB 6 PLLIELGTTE 15

RESULT 6
Y753_SVNY3 STANDARD; PRT; 449 AA.
ID Y753_SVNY3
AC P74635;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 48.0 KDA PROTEIN SLR0753.
GN SLR0753.
OS *Synechocystis* sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32739; AAC22237.1; -
DR HSSP; P13551; IELO.
DR TIGR; H10579; -
DR InterPro; IPR000640; -
DR InterPro; IPR000795; -
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR PRINTS; PR00315; ELONGATNCT.
DR PROSITE; PS00301; EFACOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding.
FT INIT_MET 0 BY SIMILARITY.
FT NP_BIND 16 23 GTP (BY SIMILARITY).
FT NP_BIND 87 91 GTP (BY SIMILARITY).
FT NP_BIND 141 144 GTP (BY SIMILARITY).
SQ SEQUENCE 699 AA; 77132 MW; FFBAD639C0F62801 CRC64;

Query Match 46.3%; Score 38; DB 1; Length 699;
Best Local Similarity 66.7%; Pred. No. 56;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 PILLLEVGVEERK 15
DB 167 PLQPLVGAEENF 178
||| ||| |||

RESULT 9
MSHL SCHPO STANDARD; PRT; 941 AA.
AC O13921; O13700;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE MUTS PROTEIN HOMOLOG 1.
GN SPAC23C11.18C OR SPAC13F5.01C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Brown D., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN MITOCHONDRIAL DNA REPAIR (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z98559; CAB11169.1; -
DR EMBL; Z99091; CAB11774.1; -

DR InterPro; IPR000432; -
DR InterPro; IPR002863; -
DR Pfam; PF00488; MTS_C; 1.
DR Pfam; PF01624; MTS_N; 1.
DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
KW DNA repair; ATP-binding; DNA-binding; Mitochondrion.
FT NP_BIND 747 754 ATP (POTENTIAL).
SQ SEQUENCE 941 AA; 106932 MW; 7A1D8F477E1140AB CRC64;

Query Match 46.3%; Score 38; DB 1; Length 941;
Best Local Similarity 53.8%; Pred. No. 76;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPELLLEVGVEEK 14
DB 60 LPPLLEKVSFQOK 72
:||||| :||

RESULT 10
PPSL_BACSU STANDARD; PRT; 2561 AA.
AC P39845;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE PEPTIDE SYNTHETASE 1.
GN PPSA OR PPS1.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95227362; PubMed=7711903;
RA Tognoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,
RA Grandi G.;
RT "A putative new peptide synthase operon in Bacillus subtilis: partial
RT characterization".
RL Microbiology 141:645-648(1995).
CC -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE
CC (POTENTIAL).
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z34883; CA84360.1; -
DR EMBL; Z99113; CAB13717.1; -
DR HSSP; P14687; IAMO.
DR Subtilist; BG10970; ppsA.
DR InterPro; IPR000255; -
DR InterPro; IPR000873; -
DR InterPro; IPR001242; -
DR Pfam; PF00501; AMP-binding; 2.
DR Pfam; PF00668; DUF4; 3.
DR Pfam; PF00550; pp-binding; 2.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
DR PROSITE; PS00455; AMP_BINDING; 2.
DR PROSITE; PS50075; ACP_DOMAIN; 2.
KW Multifunctional enzyme; Ligase; Repeat; Phosphopantetheine.
FT DOMAIN 966 1033 ACYL CARRIER (ACP).
FT DOMAIN 2012 2078 ACYL CARRIER (ACP).
FT BINDING 2042 2042 PHOSPHOPANTHETHEINE (POTENTIAL).
SQ SEQUENCE 2561 AA; 289180 MW; 5476CBE4DD882FD2 CRC64;

```

Query Match          46.3%; Score 38; DB 1; Length 2561;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 PPLLEVGVEK 15
DB 2394 PILLDMGIPDF 2405
||||:|||||

RESULT 11
REV_HVIND STANDARD; PRT; 116 AA.
AC P1803; 1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 01-FEB-1996 (Rel. 33, Last annotation update)
RE REV PROTEIN (ANTI-REPRESSION TRANSACTIVATOR PROTEIN) (ART/TRS).
GN Human immunodeficiency virus type 1 (NDK isolate) (HIV-1).
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
CC [1] TaxID=11693;
RN MEDLINE=90034200; PubMed=2806917;
RX Spire B., Sire J., Zachar V., Rey F., Barre-Sinoussi F., Calibert F.,
RA Hampe A., Chermann J.C.; HIV1-NDK: a highly cytopathic strain of the
RT Nucleotide sequence of HIV1-NDK: a highly cytopathic virus.
RL Gene 81:275-284 (1989).
CC -1- FUNCTION: REPRESSION APPEARS TO ACT POST-TRANSCRIPTIONALLY TO RELIEVE
CC NEGATIVE REPRESSION OF GAG AND ENV PRODUCTION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; ACCUMULATES IN THE NUCLEOLI.
CC -1- PM: PHOSPHORYLATION WHOSE STATE OF PHOSPHORYLATION IS REGULATED
CC BY THE NUCLEOLAR PROTEIN.
CC -1- MISCELLANEOUS: NDK ISOLATED FROM A ZAIRIAN PATIENT AFFECTED WITH
CC AIDS, AND IS A HIGHLY CYTOPATHOGENIC STRAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use. The European Bioinformatics Institute is not responsible for any
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M27323; AAA44867.1; -
CC PIR; J00072; KGLNDK
CC KEGG; 0303-2; KGLNDK
CC InterPro; IPR000525;
CC Pfam; PF00424; REV: 1.
CC Transcription regulation; AIDS; Phosphorylation; Nuclear protein.
CC SEQUENCE 116 AA; 12983 MW; 40CC1B562B655AEE CRC64;

Query Match          45.1%; Score 37; DB 1; Length 116;
Best Local Similarity 63.6%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLLEVGVEE 13
DB 106 PPAVLESGTEE 116
||||:|||||

RESULT 12
REV_HAEIN STANDARD; PRT; 139 AA.
AC P44020;
DT 01-NOV-1995 (Rel. 32, Created)
DE 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL PROTEIN HI0589.

Query Match          45.1%; Score 37; DB 1; Length 139;
Best Local Similarity 42.1%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 2; Indels 4; Gaps 1;

QY 2 VPPL---LLEVGVEEKFM 16
DB 52 ITPLRDQMVLEIGLEKSM 70
||||:|||||

RESULT 13
YRFC_ECOLI STANDARD; PRT; 179 AA.
AC P45752; 1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 20.8 KDA PROTEIN IN HOFO-MRCA INTERGENIC REGION (F179).
GN YRFC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CC [1] TaxID=562;
RN SEQUENCE FROM N.A.
RX STRAIN=K12 / MGL1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Fraser J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shach Y.
CC "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).

```

```

GN HI0589.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
CC NCBI_TaxID=727;
RN SEQUENCE FROM N.A.
RX STRAIN=K12 / MGL1655;
RX MEDLINE=94530530; PubMed=7542800;
RX Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., M.D.,
RA Weidman J.F., Phillips M.C.A., Spriggs T., Hedblom M., Branon M.D.,
RA Pinerobek K., Fritchman J.C., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512 (1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC -1- (POTENTIAL); TO E-COLI RSEC AND H-INFLUENZAE HI850.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use. The European Bioinformatics Institute is not responsible for any
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; J32740; AAC22246.1; -
CC GDB; 535; protein; 91
CC HYDROTHERMAL 71 POTENTIAL.
CC TRANSMEM 97 117
CC SEQUENCE 139 AA; 15635 MW; 6C26821FAF4DF32D CRC64;

Query Match          45.1%; Score 37; DB 1; Length 139;
Best Local Similarity 42.1%; Pred. No. 16;
Matches 8; Conservative 5; Mismatches 2; Indels 4; Gaps 1;

QY 2 VPPL---LLEVGVEEKFM 16
DB 52 ITPLRDQMVLEIGLEKSM 70
||||:|||||

RESULT 13
YRFC_ECOLI STANDARD; PRT; 179 AA.
AC P45752; 1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 20.8 KDA PROTEIN IN HOFO-MRCA INTERGENIC REGION (F179).
GN YRFC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
CC [1] TaxID=562;
RN SEQUENCE FROM N.A.
RX STRAIN=K12 / MGL1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Fraser J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shach Y.
CC "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).

```

```
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U18997; AAA58191.1; -
DR EMBL; AE000414; AAC76413.1; -
DR EcoGene; EGI2924; yrFC.
KW Hypothetical protein.
SQ SEQUENCE 179 AA; 20796 MW; 2A7D86A4B36B10C8 CRC64;

Query Match 45.1%; Score 37; DB 1; Length 179;
Best Local Similarity 63.6%; Pred. No. 20;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 YVPPLLLEGV 11
Db 26 FVAPLLLA VGI 36

RESULT 14
YACE_PSEPU
ID YACE_PSEPU STANDARD; PRT; 207 AA.
AC P36644;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL 23.0 KDA PROTEIN IN XCPA/PILD 3' REGION.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WCS358;
RA MEDLINE=94131942; PubMed=7905475;
RX de Groot A., Heijnen I., de Cock H., Filloux A., Tommassen J.;
RT "Characterization of type IV pilus genes in plant growth-promoting
RT Pseudomonas putida WCS358."
RL J. Bacteriol. 176:642-650(1994).
CC -!- SIMILARITY: BELONGS TO THE UPF0038 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X74276; GAA52335.1; -
DR PIR; S35954; S35954.
DR InterPro; IPR001977; -
DR Pfam; PF01121; UPF0038; 1.
DR PROSITE; PS01294; UPF0038; 1.
KW Hypothetical protein; ATP-binding.
FT NP_BIND 15 22 ATP (POTENTIAL).
SQ SEQUENCE 207 AA; 22975 MW; D62B74957FFED098 CRC64;

Query Match 45.1%; Score 37; DB 1; Length 207;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 YVPPLLLEGVGVEEK 14
Db 116 YVSPLLIESGVYRK 129
```

```
RESULT 15
AACT_HUMAN
ID AACT_HUMAN STANDARD; PRT; 423 AA.
AC P01011;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ALPHA-1-ANTICHYMOTRYPSIN PRECURSOR (ACT).
GN SERPINA3 OR AACT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84080367; PubMed=6606438;
RA Chandra T., Stackhouse R., Kidd V.J., Robson K.J.H., Woo S.L.C.;
RT "Sequence homology between human alpha 1-antichymotrypsin, alpha 1-
RT antitrypsin, and antithrombin III."
RL Biochemistry 22:5055-5061(1983).
RN [2]
RP SEQUENCE OF 87-129 FROM N.A.
RX MEDLINE=90110106; PubMed=2404007;
RA Rubin H., Wang Z., Nickbarg E.B., McLarney S., Naidoo N.,
RA Schoenberger O.L., Johnson J.L., Cooperman B.S.;
RT "Cloning, expression, purification, and biological activity of
RT recombinant native and variant human alpha 1-antichymotrypsins."
RL J. Biol. Chem. 265:1199-1207(1990).
RN [3]
RP SEQUENCE OF 22-423 FROM N.A.
RA Rubin H.;
RL Submitted (OCT-1989) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF N-TERMINUS.
RX MEDLINE=89323223; PubMed=2787670;
RA Lindmark B., Hilja H., Alan R., Eriksson S.;
RT "The microheterogeneity of desialylated alpha 1-antichymotrypsin: the
RT occurrence of two amino-terminal isoforms, one lacking a His-Pro
RT dipeptide."
RL Biochim. Biophys. Acta 997:90-95(1989).
RN [5]
RP ACTIVE SITE.
RX MEDLINE=84032476; PubMed=6556193;
RA Moril M., Travis J.;
RT "Amino acid sequence at the reactive site of human alpha 1-
RT antichymotrypsin."
RL J. Biol. Chem. 258:12749-12752(1983).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=91202538; PubMed=2016749;
RA Baumann U., Huber R., Bode W., Grosse D., Lesjak M., Laurell C.-B.;
RT "Crystal structure of cleaved human alpha 1-antichymotrypsin at 2.7-A
RT resolution and its comparison with other serpins."
RL J. Mol. Biol. 218:595-606(1991).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).
RX MEDLINE=96433079; PubMed=8836107;
RA Lukacs C.M., Zhong J.Q., Plotnick M.I., Rubin H., Cooperman B.S.,
RA Christianson D.W.;
RT "Arginine substitutions in the hinge region of antichymotrypsin
RT affect serpin beta-sheet rearrangement."
RL Nat. Struct. Biol. 3:888-893(1996).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE=98198038; PubMed=9521649;
RA Lukacs C.M., Rubin H., Christianson D.W.;
RT "Engineering an anion-binding cavity in antichymotrypsin modulates
RT the 'spring-loaded' serpin-protease interaction."
RL Biochemistry 37:3297-3304(1998).
RN [9]
RP VARIANTS.
RX MEDLINE=94063919; PubMed=8244391;
```


RA Piller W., Faber J.-P., Weidinger S., Tief K., Scholz S., Fischer M.,
RT Oakes M., Kimmis H., Haidmann H.-H.;
RT "A leucine-to-proline substitution causes a defective alpha 1-
RT antichymotrypsin allele associated with familial obstructive lung
RT disease.";
RL Genomics 17:740-743(1993).
RN [10]
RP VARIANT VAL-401.
RX MEDLINE=92363100; PubMed=16193300;
RY MEDLINE=92363100; Yamamura M., Yamamoto M., Shinohara Y.;
RT "Detection of a new mutant alpha-1-antichymotrypsin in patients with
RT occlusive-cerebrovascular disease.";
RL FEBS Lett. 304:66-68(1992).
CC -1- FUNCTION: ALTHOUGH ITS PHYSIOLOGICAL FUNCTION IS UNCLEAR, IT CAN
CC INHIBIT NEUTROPHIL CATHETERIN G AND SERPINS. BOTH OF
CC SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- TISSUE SPECIFICITY: PLASMA. SYNTHESIZED IN THE LIVER. LIKE
CC THE RELATED ALPHA-1-ANTITRYPSIN, ITS CONCENTRATION INCREASES IN
CC THE ACUTE PHASE OF INFLAMMATION OR INFECTION.
CC -1- DISEASE: DEFICIENCY IN ACT CAN BE A CAUSE OF CHRONIC OBSTRUCTIVE
CC PULMONARY DISEASE (COPD) OR ANEMIA. SERPINS ARE GENETICALLY
CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-4 IS THE INITIATOR.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, provided that this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; K01500; AAA51543.1;
CC EMBL; J05176; AAA51560.1;
CC PIR; A14236; A34934.
CC PIR; S14806; S14806.
CC PDB; 2ACH; 15-JUL-93.
CC PDB; 3CAA; 25-FEB-98.
CC PDB; 4CAA; 25-FEB-98.
CC PDB; 1AS4; 25-FEB-98.
CC GlycoSuiteDB; P01011;
CC KIM; S1072807; P01011; HUMAN.
CC InterPro; IPR000215;
CC Pfam; PF00079; serpin; 1.
CC PROSITE; PS00284; SERPIN; 1.
CC Serpin; Serine protease inhibitor; Plasma; Acute phase; Glycoprotein;
KW Signal; 3D-structure; Polymorphism; Disease mutation.
KW CHAIN 24.
FT ACT_SITE 383
FT CARBOHYD 93
FT CARBOHYD 127
FT DNA_BIND 235
FT VARIANT 9
FT VARIANT 78
FT VARIANT 167
FT VARIANT 252
FT VARIANT 401
FT VARIANT 69
FT CONFLICT 102
FT CONFLICT 123
FT CONFLICT 199
FT CONFLICT 361

FT CONFLICT 421
FT HELIX 48
FT TURN 69
FT STRAND 73
FT HELIX 85
FT TURN 90
FT HELIX 93
FT TURN 103
FT TURN 107
FT HELIX 112
FT STRAND 133
FT TURN 144
FT TURN 151
FT STRAND 164
FT STRAND 170
FT HELIX 173
FT TURN 188
FT TURN 201
FT STRAND 223
FT STRAND 227
FT STRAND 240
FT STRAND 257
FT STRAND 261
FT STRAND 272
FT TURN 281
FT TURN 284
FT TURN 290
FT HELIX 304
FT STRAND 316
FT STRAND 323
FT STRAND 327
FT HELIX 337
FT TURN 339
FT TURN 344
FT TURN 351
FT STRAND 365
FT STRAND 382
FT STRAND 399
FT STRAND 405
FT STRAND 414
FT TURN 421
SQ SEQUENCE 423 AA; 47650 MW; B002F946C86A8951 CRC64;
Query Match 45.1%; Score 37; DB 1; Length 423;
BLAST Local Similarity 54.5%; Pred. No. 50;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 5 LLEVGVEEKF 15
DB 327 ILLQLGIEAF 337
RESULT 16
ID ENVZ_SALTY STANDARD; PRT; 450 AA.
AC P08982; 1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last annotation update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE OSMOLARITY SENSOR PROTEIN ENVZ (EC 2.7.3.-).
GN ENVZ.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RP SEQUENCE FROM N.A.
RX STRAIN=LT2;
RX MEDLINE=89011922; PubMed=2845093;

RA Lijestrom P., Laamanen I., Palva E.T.;
 RT "Structure and expression of the ompB operon, the regulatory locus
 RT for the outer membrane porin regulon in Salmonella typhimurium
 RL LT-2";
 RL J. Mol. Biol. 201:663-673(1988).
 CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM ENVZ/OMPR
 CC INVOLVED IN THE REGULATION OF OSMOREGULATION (GENES OMPF & OMPG).
 CC ENVZ FUNCTIONS AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT
 CC PHOSPHORYLATES OMPR IN RESPONSE TO ENVIRONMENTAL SIGNALS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
 CC -!- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
 CC KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X12374; CAA30935.1; -;
 CC FIR; S01367; S01367.
 CC StyGene; SG10091; envz.
 CC InterPro; IPR000410; -;
 CC InterPro; IPR000658; -;
 CC Pfam; PF00672; DUF5; 1.
 CC Pfam; PF00512; signal; 1.
 CC Sensory transduction; Transferrase; Kinase; Phosphorylation;
 KW Transmembrane; Inner membrane.
 FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 16 35 PROBABLE.
 FT DOMAIN 36 161 PERIPLASMIC (POTENTIAL).
 FT TRANSMEM 162 182 PROBABLE.
 FT DOMAIN 183 450 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 220 450 TRANSMITTER DOMAIN (POTENTIAL).
 FT MOD_RES 243 243 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 SQ SEQUENCE 450 AA; 50331 MW; 1E424AE0CF78225B CRC64;

Query Match 45.1%; Score 37; DB 1; Length 450;
 Best Local Similarity 58.3%; Pred. No. 53;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEE 13
 :||| ||| |
 DB 201 IPPPLEYGEASE 212

RESULT 17
 PPOX_MYCLE STANDARD; PRT; 451 AA.
 AC Q50008;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROTOPORPHYRIN GEN OXIDASE (EC 1.3.3.4) (PPO).
 GN HEM.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 [1]
 RN SEQUENCE FROM N.A.
 RA Smith D.R., Robison K.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 CC -!- FUNCTION: CATALYZES THE 6-ELECTRON OXIDATION OF PROTOPORPHYRINOGEN
 CC IX TO FORM PROTOPORPHYRIN IX.
 CC -!- CATALYTIC ACTIVITY: PROTOPORPHYRINOGEN-IX + O(2) -> PROTOPORPHYRIN-
 CC IX + H(2)O(2).
 CC -!- COFACTOR: CONTAINS ONE FAD PER HOMODIMER (BY SIMILARITY).
 CC -!- PATHWAY: PENULTIMATE STEP IN HEME AND PORPHYRIN BIOSYNTHESIS.
 CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE PROTOPORPHYRINOGEN OXIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U15181; AAA62958.1; -;
 CC Porphyrin biosynthesis; Oxidoreductase; Flavoprotein; FAD;
 KW Heme biosynthesis.
 FT NP_BIND 10 15 FAD (POTENTIAL).
 SQ SEQUENCE 451 AA; 46580 MW; DF76EB1655CA2056 CRC64;

Query Match 45.1%; Score 37; DB 1; Length 451;
 Best Local Similarity 53.8%; Pred. No. 53;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEK 14
 :||| ||| |
 DB 69 VPALLAELGLSER 81

RESULT 18
 GLMS_BUCAI STANDARD; PRT; 608 AA.
 ID GLMS_BUCAI
 AC P57138;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE [ISOMERIZING]
 DE (EC 2.6.1.16) (HEXOSEPHOSPHATE AMINOTRANSFERASE) (D-FRUCTOSE-6-
 DE PHOSPHATE AMIDOTRANSFERASE) (GFAT) (L-GLUTAMINE-D-FRUCTOSE-6-PHOSPHATE
 DE AMIDOTRANSFERASE) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).
 GN GLMS OR BU026.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 OS symbiotic bacterium).
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=118099;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=TOKYO 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 RT Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC -!- FUNCTION: CATALYZES THE FIRST STEP IN HEXOSAMINE METABOLISM,
 CC CONVERTING FRUCTOSE-6P INTO GLUCOSAMINE-6P USING GLUTAMINE AS A
 CC NITROGEN SOURCE (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: L-GLUTAMINE + D-FRUCTOSE 6-PHOSPHATE ->
 CC L-GLUTAMATE + D-GLUCOSAMINE 6-PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-2
 CC GATASE DOMAIN.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.
 CC GLMS SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AP001118; BAB12753.1; ALT INIT.
 DR PROSITE; PS00443; GATASE_TYPE_II; 1.
 KW Transferrase; Aminotransferase; Glutamine amidotransferase.

```
FT INIT_MBT 0 0 BY SIMILARITY.
FT ACT_SITE 1 1 GATASE (BY SIMILARITY).
FT ACT_SITE 603 1 ISOMERIZATION PRO-6S-SEPHASE.
FT DOMAIN 183 1 GLYCINE-TRNA LIGASE.
FT SEQUENCE 608 AA: 68550 MW: 7745396CDD190BC CRC64;

Query Match 45.1%; Score 37; DB 1; Length 608;
Best Local Similarity 46.2%; Pred. No. 72;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LLLLEGVGVEKFM 16
Db 178 PLIIIGTGTEENFI 190

RESULT 19
SYGB_BACSU BACSU STANDARD; PRT; 679 AA.
AC P54381.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE GLYCYL-TRNA SYNTHETASE BETA CHAIN (EC 6.1.1.14) (GLYCINE-TRNA LIGASE
DE BETA CHAIN) (GLYS).
OS Bacillus subtilis.
OC Bacillus; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN STRAIN=168 / JH642;
RC Sato T., Takeuchi M.;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RE -1- CATALYTIC ACTIVITY: ATP + L-GLYCINE + TRNA(GLY) -> AMP +
CC PYROPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL database.
CC the European Bioinformatics Institute as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; D84432; BAA12485.1; -
CC EMBL; Z99116; CAB14453.1; -
CC SDBS; 011116; CAB14453.1; -
CC InterPro; IPR02311; 9.1.5.
CC Pfam; PF02092; CRNA_synt_2f; 1.
CC PRINTS; PR01045; TRNASYNTHGB.
CC KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
CC SEQUENCE 679 AA: 76232 MW: ACC96723A85D52BB CRC64;

Query Match 45.1%; Score 37; DB 1; Length 679;
Best Local Similarity 77.8%; Pred. No. 81;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 LLLLEGVGVEE 13
Db 6 LLLLEGLLEE 14

RESULT 20
AT7A_MOUSE STANDARD; PRT; 1491 AA.
ID AT7A_AT7A_MOUSE
```

```
AC 064430; 064431; 035101; P97422;
AC 064430; 1997 (Rel. 35, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE COPPER-TRANSPORTING ATPASE 1 (EC 3.6.3.4) (COPPER PUMP 1) (MENKES
DE DISEASE-ASSOCIATED PROTEIN HOMOLOG).
GN ATP7A OR MNK. (Mouse)
OS Mus musculus (Rodentia; Chordata; Craniata; Vertebrata; Euteleostomi;
OS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Brain;
RC MEDLINE=933244; PubMed=8054976;
RA Gitschier J., Elder B., Martin C., Verley F., Packman S.,
RA Gitschier J., Elder B., Martin C., Verley F., Packman S.,
RT "The mottled gene is the mouse homologue of the Menkes disease gene.";
RL Nat. Genet. 6:369-373(1994).
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=1G13; TISSUE=Testis;
RC MEDLINE=943214; PubMed=8054977;
RA Mercer J.F.B., Grimes A., Ambrosini L., Lockhart P., Paynter J.A.,
RA Mercer J.F.B., Grimes A., Ambrosini L., Lockhart P., Paynter J.A.,
RT "Mutations in the murine homologue of the Menkes gene in dappled and
RT blotchy mice.";
RL Nat. Genet. 6:374-378(1994).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=CBA x C3H;
RC MEDLINE=97358576; PubMed=9215672;
RA Grimes A., Hearn C.J., Lockhart P., Newgreen D.F., Mercer J.F.B.;
RA Grimes A., Hearn C.J., Lockhart P., Newgreen D.F., Mercer J.F.B.;
RT "Molecular disease": the brindled mouse mutant (No(br)); a murine model
RT Hum. Mol. Genet. 6:1037-1042(1997).
CC -1- FUNCTION: MAY SUPPLY COPPER TO COPPER-REQUIRING PROTEINS WITHIN
CC THE SECRETORY PATHWAY, WHEN LOCALIZED IN THE TRANS-GOLGI NETWORK.
CC UNDER CONDITIONS OF ELEVATED EXTRACELLULAR COPPER, IT REDUCES THE
CC TO THE PLASMA MEMBRANE WHERE IT FUNCTIONS IN THE EFFLUX OF COPPER
CC FROM CELLS (REF. 1).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + CU(2+)(IN) = ADP + PHOSPHATE +
CC CU(2+)(OUT).
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONSTITUTIVELY
CC CYCLES BETWEEN THE TRANS-GOLGI NETWORK (TGN) AND THE PLASMA
CC MEMBRANE. PREDOMINANTLY FOUND IN THE TGN AND BELONGS TO THE
CC PLASMA MEMBRANE IN RESPONSE TO ELEVATED COPPER LEVELS (BY
CC TISSUE SPECIFICITY: FOUND IN MOST TISSUES EXCEPT LIVER. IN THE
CC KIDNEY, IT IS DETECTED IN THE PROXIMAL AND DISTAL TUBULES.
CC -1- DISEASE: DEFECTS IN ATP7A ARE ASSOCIATED WITH MOTTLED, AN UNLINKED
CC RECESSIVE CONDITION CHARACTERIZED BY MORTALITY IN THE NEONATE, OR FETAL DEATH. IT
CC COAT, DEFECTS IN COPPER ABSORPTION AND TRANSPORT OF COPPER. THESE
CC MOTTLED MUTANTS EXHIBIT A DIVERSITY OF PHENOTYPES. TWO OF THESE
CC MUTANTS ARE CALLED BRINDLED AND BLOTCHY AND THEIR PHENOTYPES
CC RESEMBLE CLASSICAL MENKES DISEASE (MD) AND OCCIPITAL HORN SYNDROME
CC (OHS) IN HUMANS, RESPECTIVELY. OTHER MUTANTS ARE CALLED DAPPLED,
CC MOSAIC, TORTOISESHELL, PEWTER, ETC. SUCRINE, LEU-1478 ET LEU-1479, IS
CC MISCELLANEOUS. THE FETAL DEATH PHENOTYPE IS ASSOCIATED WITH
CC AN INCREASED FETAL SIGNAL WHICH FUNCTIONS IN RETRIEVING
CC RECYCLING FROM THE PLASMA MEMBRANE TO THE TGN. MUTATION OF THE DI-
CC LEUCINE SIGNAL RESULTS IN THE ACCUMULATION OF THE PROTEIN IN THE
```

CC PLASMA MEMBRANE (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES). SUBFAMILY 1B.
 CC -!- SIMILARITY: CONTAINS 6 HEAVY-METAL-ASSOCIATED (HMA) DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U03434; AAA57445.1; -
 DR EMBL; U03736; AAB08487.1; -
 DR EMBL; AB007134; BAA22369.1; -
 DR EMBL; U71091; AAB37301.1; -
 DR HSSP; Q04656; 2AW0.
 DR MGD; MGI:99400; Atp7a.
 DR InterPro; IPR001757; -
 DR InterPro; IPR001877; -
 DR InterPro; IPR001934; -
 DR Pfam; PF00122; E1-E2_ATPase; 2.
 DR Pfam; PF00403; HMA; 6.
 DR PRINTS; PR00119; CATATPASE.
 DR PRINTS; PR00942; CUATPASE1.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 DR PROSITE; PS01047; HMA; 6.
 KW Hydrolase; Copper transport; Transmembrane; Phosphorylation;
 KW ATP-binding; Metal-binding; Magnesium; Copper; Repeat; Golgi stack;
 KW Disease mutation; Multigene family.
 FT DOMAIN 1 644
 FT TRANSMEM 645 666
 FT DOMAIN 667 705
 FT TRANSMEM 706 725
 FT DOMAIN 726 732
 FT TRANSMEM 733 753
 FT DOMAIN 754 772
 FT TRANSMEM 773 793
 FT DOMAIN 794 826
 FT TRANSMEM 827 950
 FT DOMAIN 951 980
 FT TRANSMEM 981 1002
 FT DOMAIN 1003 1347
 FT TRANSMEM 1348 1365
 FT DOMAIN 1366 1376
 FT TRANSMEM 1377 1396
 FT DOMAIN 1397 1491
 FT DOMAIN 14 43
 FT DOMAIN 177 206
 FT DOMAIN 283 310
 FT DOMAIN 383 412
 FT DOMAIN 485 514
 FT DOMAIN 561 590
 FT DOMAIN 591 636
 FT METAL 1292 1292
 FT METAL 1296 1296
 FT SITE 1478 1478
 FT MOD_RES 1035 1035
 FT CARBOHYD 677 677
 FT CARBOHYD 966 966
 FT VARIANT 674 674
 FT VARIANT 1381 1381
 FT CONFLICT 44 44
 FT CONFLICT 103 103
 FT CONFLICT 172 172
 FT CONFLICT 245 245
 FT CONFLICT 445 445
 FT CONFLICT 470 470
 FT CONFLICT 515 515
 FT CONFLICT 717 717
 FT CONFLICT 770 770
 FT CONFLICT 775 775
 H -> R (IN MD).
 S -> P (IN MD).
 D -> E (IN REF. 2 AND 3).
 V -> I (IN REF. 2 AND 3).
 R -> M (IN REF. 2 AND 3).
 LK -> PI (IN REF. 2).
 P -> PA (IN REF. 2 AND 4).
 P -> L (IN REF. 1).
 C -> F (IN REF. 2 AND 3).
 T -> A (IN REF. 2).
 P -> S (IN REF. 2).

FT CONFLICT 885 885 I -> T (IN REF. 2).
 FT CONFLICT 1169 1169 Y -> H (IN REF. 2).
 FT CONFLICT 1204 1204 A -> P (IN REF. 2 AND 4).
 FT CONFLICT 1217 1217 I -> M (IN REF. 1).
 FT CONFLICT 1253 1253 R -> Q (IN REF. 1).
 SQ SEQUENCE 1491 AA; 161909 MW; B916EF9E2565247C CRC64;
 Query Match 45.1%; Score 37; DB 1; Length 1491;
 Best Local Similarity 53.8%; Pred. No. 1.8e+02;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 2 VPPLLEVGVEEK 14
 II I I I I I I I I
 DB 143 VPDLSDMGQTQEK 155
 RESULT 21
 ID AT7A_RAT STANDARD; PRT; 1492 AA.
 AC P70705;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COPPER-TRANSPORTING ATPASE 1 (EC 3.6.3.4) (COPPER PUMP 1) (MENKES
 DE DISEASE-ASSOCIATED PROTEIN HOMOLOGY).
 GN ATP7A OR MNK.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=101116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Astrocytes;
 RX MEDLINE=98220710; PubMed=9562241;
 RA Qian Y., Tiffany-Castiglioni E., Harris E.D.;
 RT "Sequence of a Menkes-type Cu-transporting ATPase from rat C6 glioma
 RT cells: comparison of the rat protein with other mammalian
 RT Cu-transporting ATPases.";
 RL Mol. Cell. Biochem. 181:49-61(1998).
 CC -!- FUNCTION: MAY SUPPLY COPPER TO COPPER-REQUIRING PROTEINS WITHIN
 CC THE SECRETORY PATHWAY, WHEN LOCALIZED IN THE TRANS-GOLGI NETWORK.
 CC UNDER CONDITIONS OF ELEVATED EXTRACELLULAR COPPER, IT RELOCALIZED
 CC TO THE PLASMA MEMBRANE WHERE IT FUNCTIONS IN THE EFFLUX OF COPPER
 CC FROM CELLS (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + CU(2+)(IN) = ADP + PHOSPHATE +
 CC CU(2+)(OUT).
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONSTITUTIVELY
 CC CYCLES BETWEEN THE TRANS-GOLGI NETWORK (TGN) AND THE PLASMA
 CC MEMBRANE. PREDOMINANTLY FOUND IN THE TGN AND RELOCALIZED TO THE
 CC PLASMA MEMBRANE IN RESPONSE TO ELEVATED COPPER LEVELS (BY
 CC SIMILARITY).
 CC -!- MISCELLANEOUS: THE C-TERMINAL DI-LEUCINE, LEU-1479 ET LEU-1480, IS
 CC AN ENDOCYTIC TARGETING SIGNAL WHICH FUNCTIONS IN RETRIEVING
 CC RECYCLING FROM THE PLASMA MEMBRANE TO THE TGN. MUTATION OF THE DI-
 CC LEUCINE SIGNAL RESULTS IN THE ACCUMULATION OF THE PROTEIN IN THE
 CC PLASMA MEMBRANE (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES). SUBFAMILY 1B.
 CC -!- SIMILARITY: CONTAINS 6 HEAVY-METAL-ASSOCIATED (HMA) DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U59245; AAB06393.1; -
 DR HSSP; Q04656; 2AW0.
 DR InterPro; IPR001757; -


```

RESULT 26
ID PFUF_ECOLI
INFO PFUF_ECOLI STANDARD; PRT: 262 AA.
AC P39405;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
GN PFERRIC IRON REDUCTASE PROTEIN PFUF.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RS SEQUENCE FROM N.A.55;
RX MEDLINE=95334362; PubMed=7610040;
RA Blattner F.R.; Plunkett G. III, Sofia H.J., Daniels D.L.,
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
(2) FACTORIZATION.
RN [2]
RS MEDLINE=99144129; PubMed=9950318;
RX Muller K., Matzanke B.F., Schunemann V., Trautwein A.X., Hantke K.;
RT "FhuF, an iron-regulated protein of Escherichia coli with a new type
RT of 2Fe-2S center."
CC Eur. J. Biochem. 258:1001-1008(1998).
CC -! FUNCTION: INVOLVED IN THE REDUCTION OF FERRIC IRON IN CYTOPLASMIC
CC -! PERMANENT BENDS A 2FE-2S CLUSTER.
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC; SOMETIME MEMBRANE-ASSOCIATED.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC altered and the EMBL/NCBI copyright notice is displayed. For commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC EMBL; U14003; AAC97266.1;
CC DR EMBL; AF000507; AAC977323.1;
CC DR Ecogen; EG12595; fhuF.
CC FT METAL_sulfur. 244 244 IRON-SULFUR (2FE-2S).
CC FT METAL 245 245 IRON-SULFUR (2FE-2S).
CC FT METAL 256 256 IRON-SULFUR (2FE-2S).
CC FT METAL 259 259 IRON-SULFUR (2FE-2S).
CC SQ SEQUENCE 262 AA; 30112 MW; 7708E776B1BE7783 CRC64;
Query Match 43.9%; Score 36; DB 1; Length 262;
Rest Local Similarity 53.8%; Pred. No. 45;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 2 VPPLLEVGVEEK 14
DB 104 VPPLMLALLTOEK 116
RESULT 27
PIVE_ADE40
ID PIV6_ADE40 STANDARD; PRT: 267 AA.
AC P48309;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1995 (Rel. 33, Last annotation update)
GN MINOR CAPSID PROTEIN PVI PRECURSOR.
OS Human adenovirus type 40.
OC Human adenovirus type 40.

```

```
CC -!- SIMILARITY: BELONGS TO THE GHMP KINASE FAMILY. HOMOSERINE
CC KINASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D90908; BAA17691.1; -
CC InterPro; IPR000870; -
CC InterPro; IPR001745; -
CC Pfam; PF00288; GHMP_Kinases; 1.
CC PRINTS; PR00958; HOMSERKINASE.
CC PROSITE; PS00627; GHMP_KINASES-ATP; 1.
KW Threonine biosynthesis; Transferase; Kinase; ATP-binding.
FT NP_BIND 91 101 ATP (POTENTIAL).
SQ SEQUENCE 306 AA; 32427 MW; FLDFA3B3DEE45F36 CRC64;

Query Match 43.9%; Score 36; DB 1; Length 306;
Best Local Similarity 54.5%; Pred. No. 53;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLEVGVE 12
DB 79 IPPLKLEIDLE 89

RESULT 29
MAP3_SCHPO
ID MAP3_SCHPO STANDARD; PRT; 365 AA.
AC P31397;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHEROMONE M-FACTOR RECEPTOR.
GN MAP3 OR SPAC3F10.10C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-ST711;
RX MEDLINE=93109361; PubMed=8380233;
RA Tanaka K., Davey J., Imai Y., Yamamoto M.;
RT "Schizosaccharomycetes pombe map3+ encodes the putative M-factor
receptor.";
RL Mol. Cell. Biol. 13:80-88(1993).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECEPTOR FOR THE PEPTIDE PHEROMONE M-FACTOR, A MATING
CC FACTOR OF S.POMBE. PHEROMONE SIGNALING IS ESSENTIAL FOR INITIATION
CC OF MEIOSIS IN S.POMBE. M-FACTOR SIGNALING ALONE MAY BE SUFFICIENT.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- INDUCTION: BY AT LEAST 3 TYPES OF REGULATION: THE MATING-TYPE OF
CC THE CELL, NUTRITIONAL CONDITIONS AND PHEROMONE SIGNALING.
CC -!- SIMILARITY: BELONGS TO FAMILY 4 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
CC EMBL; D10933; BAA01727.1; -
CC EMBL; Z69369; CAA93308.1; -
CC PIR; A48105; A48105.
CC GCRDB; GCR_0565; -
CC InterPro; IPR001499; -
CC Pfam; PF02076; STE3; 2.
KW Transmembrane; G-protein coupled receptor; Pheromone response.
FT TRANSMEM 7 24 POTENTIAL.
FT TRANSMEM 31 54 POTENTIAL.
FT TRANSMEM 73 100 POTENTIAL.
FT TRANSMEM 116 133 POTENTIAL.
FT TRANSMEM 155 182 POTENTIAL.
FT TRANSMEM 204 226 POTENTIAL.
FT TRANSMEM 265 283 POTENTIAL.
SQ SEQUENCE 365 AA; 42498 MW; 306A022FA7DCFEF7 CRC64;

Query Match 43.9%; Score 36; DB 1; Length 365;
Best Local Similarity 60.0%; Pred. No. 64;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YVPLLELVG 10
DB 161 YIPPCLLSFG 170

RESULT 30
DNAJ_BACSH
ID DNAJ_BACSH STANDARD; PRT; 368 AA.
AC O69269;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHAPERONE PROTEIN DNAJ.
GN DNAJ.
OS Bacillus sphaericus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1421;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=1593;
RA Ahmad S., Selvapandian A., Gasbarri M., Bhatnagar R.K.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,
CC THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).
CC -!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.
CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y17157; CAA76664.1; -
CC HSSP; P25685; 1HDJ.
CC InterPro; IPR001305; -
CC InterPro; IPR001623; -
CC InterPro; IPR002939; -
CC InterPro; IPR003095; -
CC Pfam; PF00226; DnaJ; 1.
CC Pfam; PF00684; DnaJ_CXXCXXGXG; 1.
CC Pfam; PF01556; DnaJ_C; 1.
CC PRINTS; PR00625; DNAJPROTEIN.
CC PROSITE; PS00636; DNAJ_1; 1.
CC PROSITE; PS00076; DNAJ_2; 1.
```



```

249 249 MANGANESE (BY SIMILARITY).
METAL SEQUENCE 373 AA: 33608 MW; 5EEFD786DBEEAB06 CRC64;
0:

Query Match 43.9%; Score 36; DB 1; Length 373;
Best Local Similarity 53.8%; Pred. NO. 66;
Matches 7; Conservative 3; Mismatches 0; Gaps 0;

QY 1 YVPPLLEVGVER 13
DB 74 YIVPVLGRGVDE 86
11:1

RESULT 32
EGNON DROME STANDARD; PRT: 373 AA.
ID EGNON370:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-JUL-1999 (Rel. 38, Last annotation update)
DE EMBRYONIC GONADOPHYLLIN EAGLE PROTEIN) (SPREADY PROTEIN).
DE EMBRYONIC GONADOPHYLLIN EAGLE PROTEIN) (SPREADY PROTEIN).
EGN ON KRG MEGALOGASTER (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=90059895; PubMed=2555153;
RA Rothe M., Nauber U., Jaekle H.;
RT "Three hormone receptor-like Drosophila genes encode an identical
RT DNA-binding finger.";
RL EMBO J. 8:3087-3094(1989).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=CANTON-S.
RX MEDLINE=96189342; PubMed=8625804;
RA Higashijima S., Shishido E., Matsuzaki M., Saigo K.;
RT "Bagle, a member of the steroidal hormone receptor-like gene superfamily, is
RT expressed in a subset of the Drosophila CNS.";
RL putative Pt1122-527-536(1996).
[3]
RN SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CX -1 SUBCELLULAR LOCATION: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
NR0 SUBFAMILY.
[4]
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation at
European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as content is in no way
modified and this statement is not used for advertising or commercial
entities requires a license (See http://www.isb-sib.ch/announcement/
or send an email to license@isb-sib.ch).
[5]
EMBL: X16631; CAA34626.1; -
EMBL: D43635; BAA07743.1; -
EMBL: D43634; BAA07742.1; -
PIR: S06010; S06010.
HSSP: P03372; P03372.1; -
FlyBase: FB0000560; eg.
Pfam: PF00105; zf-C4; 1.
PRINTS: PR00047; STROIDFINGER.
DR PROSITE: PS00031; NUCLEARRECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
FT Zinc-finger. 5 71 NUCLEAR RECEPTOR-TYPE.
FT ZN-FING 42 66 C4-TYPE.
FT ZN-FING 5 25 C4-TYPE.
FT SEQUENCE 373 AA; 39601 MW; 012FF997BA9A5760 CRC64;
SQ

```

```
Query Match 43.9%; Score 36; DB 1; Length 373;
Best Local Similarity 46.2%; Pred. NO. 66;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 YVPPLLEGVGVEE 13
Db 277 HLPVLLAAGVQ 289

RESULT 33
DNAJ_LISMO STANDARD; PRT; 377 AA.
ID DNAJ_LISMO STANDARD; PRT; 377 AA.
AC Q9S5A3;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHAPERONE PROTEIN DNAJ.
GN DNAJ.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=10403S;
RX MEDLINE=20163771; PubMed=10701836;
RA Hanawa T., Kai M., Kamiya S., Yamamoto T.;
RT "Cloning, sequencing, and transcriptional analysis of the dnaK heat
shock operon of Listeria monocytogenes.";
RL Cell Stress Chaperones 5:21-29(2000).
CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,
THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).
CC -!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.
CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CR DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB023064; BAA82790.1; -
CC InterPro; IPR001305; -
CC InterPro; IPR001623; -
CC InterPro; IPR002939; -
CC InterPro; IPR003095; -
CC Pfam; PF00226; DnaJ; 1.
CC Pfam; PF00684; DnaJ_CXKXG; 1.
CC Pfam; PF01556; DnaJ_C; 1.
CC PRINTS; PR00625; DNAJPROTEIN.
CC PROSITE; PS00636; DNAJ_1; 1.
CC PROSITE; PS0076; DNAJ_2; 1.
CC PROSITE; PS00637; DNAJ_CXKXG; 1.
CC Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.
CC DOMAIN 5 69
CC FT REPEAT 76 117 GLY-RICH.
CC FT REPEAT 147 154 CXXCXGXG MOTIF.
CC FT REPEAT 164 171 CXXCXGXG MOTIF.
CC FT REPEAT 190 197 CXXCXGXG MOTIF.
CC FT REPEAT 204 211 CXXCXGXG MOTIF.
CC FT METAL 147 147 ZINC 1 (BY SIMILARITY).
CC FT METAL 150 150 ZINC 1 (BY SIMILARITY).
CC FT METAL 164 164 ZINC 2 (BY SIMILARITY).
CC FT METAL 167 167 ZINC 2 (BY SIMILARITY).
CC FT METAL 190 190 ZINC 2 (BY SIMILARITY).
CC FT METAL 193 193 ZINC 2 (BY SIMILARITY).
CC FT METAL 204 204 ZINC 1 (BY SIMILARITY).
CC FT METAL 207 207 ZINC 1 (BY SIMILARITY).

SQ SEQUENCE 377 AA; 41077 MW; 70A8C5836239310A CRC64;
Query Match 43.9%; Score 36; DB 1; Length 377;
Best Local Similarity 46.7%; Pred. NO. 66;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 YVPPLLEGVGVEEK 15
Db 32 YHPDINKEAGADEKF 46

RESULT 34
DNAJ_LACLA STANDARD; PRT; 379 AA.
ID DNAJ_LACLA STANDARD; PRT; 379 AA.
AC P35514;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHAPERONE PROTEIN DNAJ.
GN DNAJ.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIZO R5;
RX MEDLINE=93194788; PubMed=8449872;
RA van Asseldonk M., Simons A., Visser H., de Vos W.M., Simons G.;
RT "Cloning, nucleotide sequence, and regulatory analysis of the
Lactococcus lactis dnaJ gene.";
RL J. Bacteriol. 175:1637-1644(1993).
CC -!- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRPE,
THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).
CC -!- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE DNAJ FAMILY.
CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 CR DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M99413; -; NOT_ANNOTATED_CDS.
CC PIR; A47079; A47079.
CC HSP; P08622; 1XBL.
CC InterPro; IPR001305; -
CC InterPro; IPR001623; -
CC InterPro; IPR002939; -
CC InterPro; IPR003095; -
CC Pfam; PF00226; DnaJ; 1.
CC Pfam; PF01556; DnaJ_C; 1.
CC Pfam; PF00684; DnaJ_CXKXG; 1.
CC PRINTS; PR00625; DNAJPROTEIN.
CC PROSITE; PS00636; DNAJ_1; 1.
CC PROSITE; PS0076; DNAJ_2; 1.
CC PROSITE; PS00637; DNAJ_CXKXG; 1.
CC Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding.
CC DOMAIN 5 69
CC FT REPEAT 72 113 GLY-RICH.
CC FT REPEAT 154 161 CXXCXGXG MOTIF.
CC FT REPEAT 171 178 CXXCXGXG MOTIF.
CC FT REPEAT 197 204 CXXCXGXG MOTIF.
CC FT REPEAT 211 218 CXXCXGXG MOTIF.
CC FT METAL 154 154 ZINC 1 (BY SIMILARITY).
CC FT METAL 157 157 ZINC 1 (BY SIMILARITY).
CC FT METAL 171 171 ZINC 2 (BY SIMILARITY).
```

NCBI_Taxid=552;
[1] - SPKAK HUMAN
SEQUENCE FROM N.A.
STRAIN=EA7/74;
Geier G., Geider K.K.;
"Characterization and influence on virulence of the levansucrase gene
from the firebrill pathogen *Erwinia amylovora*,"
Physiol. Mol. Plant Path. 48: 99-104, 1993.
-!- GLUCOSE + (2,6-BETA-D-FRUCTOSYL) (N) =
GLUCOSE + (2,6-BETA-D-FRUCTOSYL) (N+1) (OTHER SUGARS CAN
ALSO ACT AS FRUCTOSYL ACCEPTORS).
-!- SUBCELLULAR LOCATION: SECRETED.
-!- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
Geneva. The Swiss Institute of Bioinformatics has no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announcement/
or send an email to licensed@isb-sib.ch).
EMBL: X75079; CAA52972.1; SwissProt: P12542.
TRANSLATION: 415 AA. 46408 MW: 4FF564F6E0607FEB CRC64;
SEQUENCE 415 AA. 46408 MW: 4FF564F6E0607FEB CRC64;
Query Match 43.9%; Score 36; DB 1; Length 415;
Best Local Similarity 46.2%; Pred. No. 73; 3; Indels 0; Gaps 0;
Matches 6; Conservative
QY 2 VPPLLEGVGVEEK 14
DB 273 LPPLITAVGVNDQ 285
RESULT HUMAN
ID SPKAK HUMAN STANDARD; PRT: 547 AA.
CD CD000001 QUERA4: 014774;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE STE20/SPS1-RELATED PROLINE-ALANINE RICH PROTEIN KINASE (BC 2.7.1.-)
DE (STE-20 RELATED KINASE) (DCHT).
GN SPKAK.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_Taxid=9606;
[1]
RN SEQUENCE: FROM N.A.
RP TISSUE=Brain.
RC MEDLINE=20438689; PubMed=10980603;
RX DOIS=10.1093/oxfordjournals.ajph.a1000001;
RA Gotoh Y., Hatakeyama H., Nishitani G., Gotoh Y., Harrison L.C.,
RT "SPAK, a STE20/SPS1-related kinase that activates the p38 pathway,"
RL Oncogene 19: 4290-4297 (2000).
[2]
RN SEQUENCE FROM N.A.
RP Melnick M.B., Pettit M., Perrimon N., Comb M.J.;
RA New human members of the Ste20 family,"
RT Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE OF 281-547 FROM N.A.
RP TISSUE=Testis;
RC Baytel D., Don J.;
RA Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: MAY ACT AS A MEDIATOR OF STRESS-INDUCED
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (WHEN CASPASE-
CC ACTIVATION SPECIFICITY). PREDOMINANTLY EXPRESSED IN BRAIN AND PANCREAS
CC -!- TISSUE SPECIFICITY. LUNG, KIDNEY, SKELETAL MUSCLE, LIVER, PLACENTA
CC FOLLOWED BY HEART.

```

CC AND TESTIS.
CC -!- DOMAIN: PAPA BOX (PROLINE-ALANINE REPEATS) MAY TARGET THE KINASE
CC TO A SPECIFIC SUBCELLULAR LOCATION BY FACILITATING INTERACTION
CC WITH INTRACELLULAR PROTEINS SUCH AS ACTIN OR ACTIN-LIKE PROTEINS.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF099989; AAC72238.1; -
DR EMBL; AF030403; AAD01901.1; -
DR EMBL; AF017635; AAB70552.1; -
DR HSSP; P00518; LPHK.
DR InterPro; IPR000719; -
DR Pfam; PF00069; pkinase; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 14 55 PRO/ALA-RICH.
FT NP_BIND 65 339 PROTEIN KINASE.
FT BINDING 71 79 ATP (BY SIMILARITY).
FT ACT_SITE 94 94 ATP (BY SIMILARITY).
FT DOMAIN 194 194 BY SIMILARITY.
FT SITE 362 368 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 389 393 CASPASE CLEAVAGE RELATED SITE.
FT CONFLICT 175 175 L -> F (IN REF. 1).
SQ SEQUENCE 547 AA; 59642 MW; DELC71CD46A79986 CRC64;

Query Match 43.9%; Score 36; DB 1; Length 547;
Best Local Similarity 61.5%; Pred. No. 98;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 4 PLLLEGVGVEEKPM 16
Db 292 PPTLETGVDEKEM 304

RESULT 38
SPAK_RAT
ID SPK_RAT STANDARD; PRT; 553 AA.
AC O88506; O70541;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE STE20/SPSI-RELATED PROLINE-ALANINE RICH PROTEIN KINASE (BC 2.7.1.-)
DE (STE-20 RELATED KINASE) (PANCREATIC SERINE/THREONINE KINASE) (PS/TK)
DE (PSTK1).
GN SPK OR PASK.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98342058; PubMed=9675032;
RA Ushiro H., Tsutsumi T., Suzuki K., Kayahara T., Nakano K.;
RT "Molecular cloning and characterization of a novel Ste20-related
RL protein kinase enriched in neurons and transporting epithelia.";
RL Arch. Biochem. Biophys. 355:233-240(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Pancreas;
RX MEDLINE=20445994; PubMed=10990492;
RA Miao N., Fung B., Sanchez R., Lydon J., Barker D., Pang K.;

```

```

RT RT Isolation and expression of PASK, a serine/threonine kinase, during
RT rat embryonic development, with special emphasis on the pancreas.";
RL J. Histochem. Cytochem. 48:1391-1400(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Insulinoma;
RX MEDLINE=20438689; PubMed=10980603;
RA Johnston A.M., Naselli G., Genez L.J., Martin R.M., Harrison L.C.,
RA de Aizpurua H.J.;
RT "SPAK, a STE20/SPSI-related kinase that activates the p38 pathway.";
RL Oncogene 19:4290-4297(2000).
CC -!- FUNCTION: MAY ACT AS A MEDIATOR OF STRESS-ACTIVATED SIGNALS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (WHEN CASPASE-
CC CLEAVED) (PROBABLE).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTES FOLLOWED BY
CC PANCREAS, KIDNEY, HEART AND BRAIN. NOT EXPRESSED IN SKELETAL
CC MUSCLE, LIVER, LUNG AND SPLEEN.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN THE EARLY GUT AND PANCREATIC
CC EPITHELIUM, AT E15 DAY LOCALIZED TO CELLS THAT WILL EVENTUALLY
CC BECOME EXOCRINE. EXPRESSED IN CHOROID PLEXUS, DEVELOPING
CC MYOCARDIUM, PANCREATIC EPITHELIUM AND DORSAL ROOT GANGLIA.
CC -!- DOMAIN: PAPA BOX (PROLINE-ALANINE REPEATS) MAY TARGET THE KINASE
CC TO A SPECIFIC SUBCELLULAR LOCATION BY FACILITATING INTERACTION
CC WITH INTRACELLULAR PROTEINS SUCH AS ACTIN OR ACTIN-LIKE PROTEINS.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D88190; BAA26000.1; -
DR EMBL; AF068261; AAC23501.1; -
DR EMBL; AF099990; AAC72239.1; -
DR HSSP; P00518; LPHK.
DR InterPro; IPR000719; -
DR Pfam; PF00069; pkinase; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 14 62 PRO/ALA-RICH.
FT NP_BIND 72 346 PROTEIN KINASE.
FT BINDING 78 86 ATP (BY SIMILARITY).
FT ACT_SITE 101 101 ATP (BY SIMILARITY).
FT DOMAIN 201 201 BY SIMILARITY.
FT DOMAIN 22 30 POLY-ALA.
FT DOMAIN 369 375 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT SITE 396 400 CASPASE CLEAVAGE RELATED SITE.
FT CONFLICT 11 11 V -> I (IN REF. 2).
FT CONFLICT 253 253 M -> I (IN REF. 2).
FT CONFLICT 403 403 E -> Q (IN REF. 2).
SQ SEQUENCE 553 AA; 60050 MW; 8BC1458AEA6F90C8 CRC64;

Query Match 43.9%; Score 36; DB 1; Length 553;
Best Local Similarity 61.5%; Pred. No. 99;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 4 PLLLEGVGVEEKPM 16
Db 299 PPTLETGVDEKEM 311

RESULT 39
SPAK_MOUSE
ID SPK_MOUSE STANDARD; PRT; 556 AA.
AC O921W9;
DT 01-OCT-2000 (Rel. 40, Created)

```

Plasmid IncFI ColV3-K30.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
NCBI_TaxID=562;

[1]
SEQUENCE FROM N.A.

CC STRAIN=K12
CX MEDLINE=94210503; PubMed=8003107;
OC MEDLINE=94210503; PubMed=8003107;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

RT "The organization of intercalstrophic regions of the aerobactin operon
of pColV-K30 may account for the differential expression of the
lucABCD iuta genes.";
RL J. Mol. Biol. 238:288-293(1994).
RC -!- PATHWAY: AEROBACTIN SIDEROPHORE BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE IUCF FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to licenses@isb-sib.ch).

DR EMBL: X76100; CAA53709.1 ;
KW Iron transport; Plasmid.
SQ SEQUENCE 580 AA; 66845 MW; 406FB857BB367F65 CRC64;

Query Match 43.9%; Score 36; DB 1; Length 580;
Best local Similarity 35.7%; Prod No. 1e+02;
Matches 3; Conservative 6; Mismatches 3; Indels 0; Gaps 0

OY 1 YVPPLLEVGVEEK 14
Db 491 FISPMLVLGVPER 504
:: :::: :||| :

RESULT 41
GLMS_CULPN STANDARD; PRT: 608 AA.
ID GLMS_CULPN
AC Q9Z6UO; O9JRV0;
DC 30-MAY-2000 (Rel. 39, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GLUCOSAMINE-6-FRUCTOSE-6-PHOSPHATE AMINO TRANSFERASE [ISOMERING] (
DE PHOSPHATE AMIDO TRANSFERASE) (GFAT) (L-GLUTAMINE-D-FRUCTOSE-6-
DE AMIDO TRANSFERASE) (GLUCOSAMINE-6-PHOSPHATE SYNTHASE).
GN GLMS OR CPN0968 OR CP0892.
OS Chlamydia pneumoniae (Chlamydiophila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydozoellales.
OX NCBI_TaxID=83558;
RN [1] DEDUCE FROM N.A.
RP STRAIN=CWL029;
RX MEDLINE=992086606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.; "Comparative
RT genomes of Chlamydia pneumoniae and C. trachomatis."; Nat. Genet. 21:385-389(1999).
RL
RN
RN SEQUENCE FROM N.A.
KC STRAIN=AR39;
EX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brumham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Umayan L.A., Utterback T.,
RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA Bowman C., Dodson R., Gwinn Miller Nelson W., DeLong E.M.,
RA McClarty G., Salzberg S.L.; "Genomic evidence of Chlamydia trachomatis MoPn and Chlamydia
TR genome lineage AR39".
RL Nucleic Acids Res. 28:1397-1406(2000).

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- FUNCTION: CATALYZES THE FIRST STEP IN HEXOSAMINE METABOLISM,
CC CONVERTING FRUCTOSE-6P INTO GLUCOSAMINE-6P USING GLUTAMINE AS A
CC NITROGEN SOURCE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-GLUTAMINE + D-FRUCTOSE 6-PHOSPHATE =
CC L-GLUTAMATE + D-GLUCOSAMINE 6-PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-2
CC GATASE DOMAIN.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.
CC GLMS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE001677; AAD19104.1; -
DR EMBL; AE002247; AAF38679.1; -
DR EMBL; AP02548; BAA99176.1; -
DR TIGR; CP0892; -
DR MEROPS; C44.971; -
DR InterPro; IPR000583; -
DR InterPro; IPR001347; -
DR Pfam; PF00310; SIS; 2.
DR PROSITE; PS00443; GATASE_TYPE_II; 1.
KW Transferase; Aminotransferase; Glutamine amidotransferase.
FT INIT_MET 0 0 BY SIMILARITY.
FT ACT_SITE 1 1 GATASE (BY SIMILARITY).
FT ACT_SITE 603 603 ISOMERISATION FRU-6P (BY SIMILARITY).
FT DOMAIN 1 182 GLUTAMINE AMIDOTRANSFERASE.
FT CONFLICT 320 320 I -> T (IN REF. 1).
SQ SEQUENCE 608 AA; 67071 MW; F274760BDBF54ED CRC64;

Query Match 43.9%; Score 36; DB 1; Length 608;
Best Local Similarity 53.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 PLLLEVGVEKFM 16
DB 178 PLLILGLKEETFI 190

RESULT 42
EFG_ECOLI STANDARD; PRT; 703 AA.
AC P02996;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ELONGATION FACTOR G (EF-G).
GN FUS OR FUS OR FAR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84144098; PubMed=6322136;
RA Zengel J.M., Archer R.H., Lindahl L.;

```

```

RT "The nucleotide sequence of the Escherichia coli fus gene, coding for
RT elongation factor G.";
RL Nucleic Acids Res. 12:2181-2192(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93013025; PubMed=1398129;
RA Johanson U., Hughes D.;
RT "Comparison of the complete sequence of the str operon in Salmonella
RT typhimurium and Escherichia coli.";
RL Gene 120:93-98(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE OF 1-93 FROM N.A.
RX MEDLINE=80182129; PubMed=6989816;
RA Post L.E., Nomura M.;
RT "DNA sequences from the str operon of Escherichia coli.";
RL J. Biol. Chem. 255:4660-4666(1980).
RN [5]
RP SEQUENCE.
RX MEDLINE=82187171; PubMed=7042386;
RA Ovchinnikov Y.A., Alakhov Y.B., Bundulis Y.P., Bundule M.A.,
RA Dovgas N.V., Kozlov V.P., Motuz L.P., Vinokurov L.M.;
RT "The primary structure of elongation factor G from Escherichia coli.
RT A complete amino acid sequence.";
RL FEBS Lett. 139:130-135(1982).
RN [6]
RP SEQUENCE OF 476-704.
RX MEDLINE=81212798; PubMed=7016587;
RA Alakhov Y.B., Dovgas N.V., Motuz L.P., Vinokurov L.M.,
RA Ovchinnikov Y.A.;
RT "The primary structure of the elongation factor G from Escherichia
RT coli: amino acid sequence of the C-terminal domain.";
RL FEBS Lett. 126:183-186(1981).
RN [7]
RP SEQUENCE OF 683-704 FROM N.A.
RX MEDLINE=81165557; PubMed=7011903;
RA Yokota T., Sugisaki H., Takanami M., Kaziro Y.;
RT "The nucleotide sequence of the cloned tufA gene of Escherichia
RT coli.";
RL Gene 12:25-31(1980).
RN [8]
RP SEQUENCE OF 1-20 FROM N.A.
RC STRAIN=L44;
RA Weigel C.T.O.;
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE OF 1-12.
RC STRAIN=K12 / EMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```
RESULT 44
DPOL_HAEIN
ID DPOL_HAEIN STANDARD; PRT; 930 AA.
AC P43741;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE DNA POLYMERASE I (EC 2.7.7.7) (POL I).
GN POLA OR HI0856.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kleravage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RT Science 269:496-512(1995).
CC -!- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC N PYROPHOSPHATE + DNA(N).
CC -!- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS.
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32767; AAC22515.1; -
DR HSSP; P00582; 1KRP.
DR TIGR; HI0856; -
DR InterPro; IPR001098; -
DR InterPro; IPR002298; -
DR InterPro; IPR002421; -
DR InterPro; IPR002562; -
DR Pfam; PF01612; 3_5-exonuclease; 1.
DR Pfam; PF01367; 5_3-exonuclease; 1.
DR Pfam; PF00476; DNA_pol_A; 1.
DR PRINTS; PR00868; DNAPOLI.
DR PROSITE; PS00447; DNA_POLYMERASE_A; 1.
KW Transferrase; DNA-directed DNA polymerase; DNA replication; DNA repair;
KW Hydrolase; Exonuclease; DNA-binding.
SQ SEQUENCE 930 AA; 103740 MW; 226654BB7CFF730B CRC64;

Query Match 43.9%; Score 36; DB 1; Length 930;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEK 15
DB 913 VVPLIVEGVGQNW 926

RESULT 45
INVA_YERPS
```

```
ID INVA_YERPS STANDARD; PRT; 985 AA.
AC P11922;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE INVASIN.
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87301720; PubMed=3304658;
RA Isberg R.R., Voorhis D.L., Falkow S.;
RT "Identification of invasin: a protein that allows enteric bacteria to
RT penetrate cultured mammalian cells.";
RL Cell 50:769-778(1987).
RN [2]
RP INTEGRIN-BINDING DOMAIN.
RX MEDLINE=90269235; PubMed=1693333;
RA Leong J.M., Fournier R.S., Isberg R.R.;
RT "Identification of the integrin binding domain of the Yersinia
RT pseudotuberculosis invasin protein.";
RL EMBO J. 9:1979-1989(1990).
CC -!- FUNCTION: INVASIN IS A PROTEIN THAT ALLOWS ENTERIC BACTERIA TO
CC PENETRATE CULTURED MAMMALIAN CELLS. THE ENTRY OF INVASIN IN THE
CC CELL IS MEDIATED BY BINDING SEVERAL BETA-1 CHAIN INTEGRINS.
CC -!- SUBCELLULAR LOCATION: OUTER SURFACE.
CC -!- SIMILARITY: BELONGS TO THE EAE/INVASIN FAMILY.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-17 OR MET-19 IS THE
CC INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17448; AAA27633.1; -
DR EMBL; M17448; AAA27632.1; ALT INIT.
DR EMBL; M17448; AAA27634.1; ALT INIT.
DR EMBL; M17448; AAA27635.1; ALT_INIT.
DR PIR; A29646; A29646.
DR PIR; S12543; S12543.
FT DOMAIN 795 985 INTEGRIN-BINDING.
SQ SEQUENCE 985 AA; 106627 MW; C875941B24BD35EE CRC64;

Query Match 43.9%; Score 36; DB 1; Length 985;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLEVGVEEK 14
DB 320 YTPVPLLTGVGDQR 333

Search completed: June 28, 2001, 11:57:22
Job time: 263 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:52:59 ; Search time 72.61 Seconds
(without alignments)
29.154 Million cell updates/sec

Title: US-09-439-313-554
Perfect score: 82
Sequence: 1 YVPPILLLEVGEVEKFM 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_unclassified.*
- 13: sp_vertebrate.*
- 14: sp_virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	59.8	599	5 Q9VSV1	Q9vsv1 drosophila
2	46	56.1	280	2 Q9RKH8	Q9rk88 streptomyc
3	45	54.9	299	2 Q56171	Q56171 streptomyc
4	43	52.4	351	10 Q9LSL1	Q9ls11 arabidopsis
5	42	51.2	230	1 Q9HL29	Q9hl29 thermoplasm
6	42	51.2	316	2 Q9FC55	Q9fc55 streptomyc
7	42	51.2	367	2 Q9WY66	Q9wy66 thermotoga
8	42	51.2	449	2 Q9X3R1	Q9x3r1 pseudomonas
9	42	51.2	528	5 Q16518	Q16518 caenorhabdi
10	42	51.2	700	2 Q91338	Q91338 pseudomonas
11	41	50.0	260	2 Q54386	Q54386 streptomyc
12	41	50.0	309	2 Q91217	Q91217 pseudomonas
13	41	50.0	405	2 Q9X6R2	Q9x6r2 pseudomonas
14	41	50.0	436	2 Q67814	Q67814 aquifex aeo
15	41	50.0	3472	1 Q74056	Q74056 cenarchaeum
16	40	48.8	108	3 Q08504	Q08504 saccharomyc
17	40	48.8	136	2 Q56530	Q56530 unidentified
18	40	48.8	157	4 Q9NW52	Q9nw52 homo sapien
19	40	48.8	243	1 Q28678	Q28678 archaeoglob

20	40	48.8	261	5 Q9GVK5	Q9gvk5 leishmania
21	40	48.8	348	1 Q9Y8P8	Q9y8p8 aeropyrum p
22	40	48.8	407	5 Q9N389	Q9n389 caenorhabdi
23	40	48.8	437	10 Q96544	Q96544 beta vulgar
24	40	48.8	687	2 Q9JXQ5	Q9jxq5 neisseria m
25	40	48.8	687	2 Q9JW67	Q9jw67 neisseria m
26	40	48.8	737	2 Q51274	Q51274 borrelia bu
27	40	48.8	916	2 Q9LCT3	Q9lct3 pseudomonas
28	39	47.6	162	2 Q9K5S0	Q9k5s0 bacillus ha
29	39	47.6	286	5 Q9UOL8	Q9uol8 plasmodium
30	39	47.6	299	4 Q9S685	Q9s685 homo sapien
31	39	47.6	307	2 P72208	P72208 prochloroco
32	39	47.6	326	2 Q9KXR2	Q9kxr2 streptomyc
33	39	47.6	333	3 Q9P727	Q9p727 neosporea
34	39	47.6	419	5 P91430	P91430 caenorhabdi
35	39	47.6	531	5 Q21603	Q21603 caenorhabdi
36	39	47.6	570	10 Q9XEE4	Q9xee4 arabidopsis
37	39	47.6	606	2 Q9PLA4	Q9pla4 chlamydia m
38	39	47.6	674	4 Q9H6U0	Q9h6u0 homo sapien
39	39	47.6	800	5 Q9XWL9	Q9xwl9 caenorhabdi
40	39	47.6	1148	4 Q9H6W7	Q9h6w7 homo sapien
41	39	47.6	1663	10 Q9LXT9	Q9lxt9 arabidopsis
42	39	47.6	1972	10 Q9LUD7	Q9lud7 arabidopsis
43	38	46.3	227	2 Q9HV03	Q9hv03 pseudomonas
44	38	46.3	259	2 Q86303	Q86303 streptomyc
45	38	46.3	277	2 Q9ZBW7	Q9zbw7 streptomyc
46	38	46.3	371	2 Q9L8R0	Q9l8r0 pseudomonas
47	38	46.3	387	2 Q9XON1	Q9xon1 thermotoga
48	38	46.3	417	2 Q9XC39	Q9xc39 pseudomonas
49	38	46.3	419	5 Q44549	Q44549 caenorhabdi
50	38	46.3	460	14 Q65854	Q65854 beet yellow
51	38	46.3	490	5 Q18169	Q18169 caenorhabdi
52	38	46.3	561	1 Q9V2L7	Q9v2l7 pyrococcus
53	38	46.3	635	2 Q9X283	Q9x283 thermotoga
54	38	46.3	1049	2 Q31100	Q31100 pseudomonas
55	38	46.3	1408	10 Q9FWU0	Q9fwu0 oryza sativ
56	38	46.3	1930	10 Q9SHJ3	Q9shj3 arabidopsis
57	38	46.3	2140	3 Q02316	Q02316 lentinula e
58	38	46.3	2581	5 Q9N4C8	Q9n4c8 caenorhabdi
59	38	46.3	4564	5 Q77075	Q77075 drosophila
60	37	45.1	164	2 Q9K5R8	Q9k5r8 bacillus ha
61	37	45.1	196	2 Q9EVK0	Q9evk0 sinorhizobi
62	37	45.1	207	2 Q69082	Q69082 pseudomonas
63	37	45.1	225	1 Q29453	Q29453 archaeoglob
64	37	45.1	266	2 Q9X8J4	Q9x8j4 streptomyc
65	37	45.1	288	10 Q9ZQB0	Q9zqb0 arabidopsis

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	599 AA.
Q9VSV1				
ID	Q9VSV1			
AC	Q9VSV1			
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)			
DE	CG4484 PROTEIN.			
GN	CG4484			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BERKELEY;			
RX	MEDLINE-20196006; PubMed-10731132;			
RA	Adams M.D., Celisliker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			

RESULT 4
Q9LSL1 PRELIMINARY; PRT; 351 AA.
AC Q9LSL1;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 5, TAC CLONE:K21L13.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB026639; BAA98183.1;
DR InterPro: IPR000515;
DR InterPro: IPR001092;
DR InterPro: IPR003015;
DR Pfam: PF00010; HLH; 1.
DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR SMART: SM00353; HLH; 1.
SQ SEQUENCE 351 AA; 39484 MW; A28579226445A59 CRC64;

Query Match 52.4%; Score 43; DB 10; Length 351;
Best Local Similarity 61.5%; Pred. No. 31;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 PPLLEVGVEEKF 15
DB 124 PPLILQGEENF 136
||||:|||||

RESULT 5
Q9HLZ9 PRELIMINARY; PRT; 230 AA.
AC Q9HLZ9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PROTEASE (PSPA) RELATED PROTEIN.
GN TA0074.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
acidophilum.";
RL Nature 407:508-513(2000).
DR EMBL: AL445063; CAC11222.1;
DR InterPro: IPR002142;
DR Pfam: PF01343; Peptidase_D7; 1.
SQ SEQUENCE 230 AA; 26584 MW; 6C60B3F66D952DA4 CRC64;

Query Match 51.2%; Score 42; DB 1; Length 230;
Best Local Similarity 57.1%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 YVPLLEVGVEEK 14
DB 107 YIRPLDKIGVENK 120
||||:|||||
RESULT 6
Q9FC55 PRELIMINARY; PRT; 316 AA.
AC Q9FC55;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PUTATIVE LIPASE.
GN SC4B10.32.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL391515; CAC04237.1;
SQ SEQUENCE 316 AA; 34103 MW; E9E20D2C8766549A CRC64;

Query Match 51.2%; Score 42; DB 2; Length 316;
Best Local Similarity 80.0%; Pred. No. 41;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 PPLLEVGVE 12
DB 20 PPLLYELGVE 29
||||:|||||

RESULT 7
Q9WYC6 PRELIMINARY; PRT; 367 AA.
AC Q9WYC6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN TM0290.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Raft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from


```

OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004596; AAG05078.1; -.
KW Hypothetical protein.
SQ SEQUENCE 700 AA; 77984 MW; 08506AA9CB310FB4 CRC64;

Query Match 51.2%; Score 42; DB 2; Length 700;
Best Local Similarity 57.1%; Pred. No. 95;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEKF 15
Db 559 VPPLLIAPGQIEKF 572

RESULT 11
Q54386
ID Q54386 PRELIMINARY; PRT; 260 AA.
AC Q54386;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE LINCOMYCIN RESISTANCE METHYLASE.
GN LRM.
OS Streptomyces lividans.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92104506; PubMed=1761231;
RA Jenkins G., Cundliffe E.;
RT "Cloning and characterization of two genes from Streptomyces lividans
RT that confer inducible resistance to lincomycin and macrolide
RT antibiotics.";
RL Gene 108:55-62(1991).
DR EMBL: M74717; AAA26779.1; -.
DR InterPro; IPR000051; -.
DR InterPro; IPR001737; -.
DR Pfam; PF00398; RnaAD; 1.
DR PROSITE; PS01131; RNA_A_DIMETH; 1.
KW Methyltransferase.
SQ SEQUENCE 260 AA; 28837 MW; 83345F6452DF2215 CRC64;

Query Match 50.0%; Score 41; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PPLLEVG 10
Db 36 PPLLEVG 43

RESULT 12
Q91217
ID Q91217 PRELIMINARY; PRT; 309 AA.
AC Q91217;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

```

```

DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PROBABLE ESTERASE/DEACETYLASE.
GN PA2098.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004637; AAG05486.1; -.
DR InterPro; IPR00379; -.
DR InterPro; IPR002168; -.
DR PROSITE; PS01173; LIPASE_GDXG_HIS; 1.
DR PROSITE; PS01174; LIPASE_GDXG_SER; UNKNOWN_1.
SQ SEQUENCE 309 AA; 33018 MW; 69CAFF3CB0F9CD33 CRC64;

Query Match 50.0%; Score 41; DB 2; Length 309;
Best Local Similarity 58.3%; Pred. No. 59;
Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEE 13
Db 232 LPPLLVQVGEDE 243

RESULT 13
Q9X6R2
ID Q9X6R2 PRELIMINARY; PRT; 405 AA.
AC Q9X6R2;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HI03;
RA Adewoye L.O., Worobec E.A.;
RT "Identification of a trans-regulatory locus involved in the regulation
RT of carbohydrate transport in Pseudomonas aeruginosa.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF143948; AAD32694.1; -.
SQ SEQUENCE 405 AA; 44968 MW; BF441FA50E1B7D58 CRC64;

Query Match 50.0%; Score 41; DB 2; Length 405;
Best Local Similarity 57.1%; Pred. No. 78;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 VPPLLEVGVEEKF 15
Db 264 VPPLLIQFGIQEKF 277

RESULT 14
Q67814
ID Q67814 PRELIMINARY; PRT; 436 AA.
AC Q67814;

```

Thu Jun 28 12:05:44 2001

```

DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DE PROTOPORPHYRINOGEN OXIDASE.
GN HEMG.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham R.A., Overbeek R., Sneed M.A., Keller M., Aubay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G., et al.
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RL Nature 392:353-358(1998).
DR EMBL: AE000768; AAC07778.1; -.
DR InterPro: IPR000205; -.
DR DR InterPro: IPR002937; -.
DR DR Pfam: PF01593; Amino-oxidase; 1.
SQ SEQUENCE 436 AA; 4987 MW; ED5F2B1BCD1DEE7 CRC64;

Query Match 50.0%; Score 41; DB 2; Length 436;
Best Local Similarity 53.3%; Pred. No. 85;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy 1 YVPLLLEVGVEEF 15
Db 279 YPVPVVVNVGVGEKF 293

RESULT 15
ID 074056 PRELIMINARY; PRT; 3472 AA.
AC Q56530;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 367.1 kDa PROTEIN.
OS Cenarchaeum symbiosum;
OC Euryarchaeota; Cenarchaeum.
OX NCBI_TaxID=46770;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=B; PubMed=9748430;
RX Schleper C., DeLong E.F., Preston C.M., Feldman R.A., Wu K.Y.,
RT "Genomic analysis reveals chromosomal variation in natural populations
RT of the uncultured psychrophilic archaeon Cenarchaeum symbiosum.";
RL J. Bacteriol. 180:5003-5009(1998).
DR EMBL: AF083072; AAC62699.1; -.
DR InterPro: IPR000515; -.
DR DR InterPro: IPR000430; 5.
DR DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.
DR SMART: SM00320; WD40; 1.
DR Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 3472 AA; 367061 MW; 37F80707030F9355 CRC64;

Query Match 50.0%; Score 41; DB 1; Length 3472;
Best Local Similarity 66.7%; Pred. No. 7.8e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 2 VPPLLLEVGVEE 13
Db 1169 LPPLFLGVGAEE 1180

```

```

RESULT 16
ID Q08504 PRELIMINARY; PRT; 108 AA.
AC Q56530;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ORF Y0R105W.
GN Y0R105W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC NCBI_TaxID=4932;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=V;
RX Voss H., Benes V., Rechmann S., Teodoru C., Schwager C., Paces V.,
RA Ansoerge W.; (JUL-1996) to the EMBL/GenBank/DBJ databases.
RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2] SEQUENCE FROM N.A.
RC MIPS;
RX Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL: Z75012; CA0300.1; -.
DR SCD; SCD; X08105M.
SQ SEQUENCE 108 AA; 12335 MW; F89561B3720B66A5 CRC64;

Query Match 48.8%; Score 40; DB 3; Length 108;
Best Local Similarity 40.0%; Pred. No. 6;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Oy 2 VPPLLLEVGVEEF 16
Db 69 LPPLIEVHIQPKII 83

RESULT 17
ID Q56530 PRELIMINARY; PRT; 136 AA.
AC Q56530;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 05, Last annotation update)
DE NITROGENASE SUBUNIT OF DINITROGENASE (MOFE PROTEIN) (FRAGMENT).
OS unidentified nitrogen-fixing bacteria.
OC Bacteria.
OX NCBI_TaxID=34107;
RN [1] SEQUENCE FROM N.A.
RC Ueda T.;
RX Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
RN [2] SEQUENCE FROM N.A.
RX MEDLINE=95254398; PubMed=7736355;
RX Ueda T., Suga Y., Yahiro N., Matsuguchi T.;
RT "Genetic diversity of N2-fixing bacteria associated with rice roots by
RT molecular analysis of N2-fixing bacteria associated with rice roots by
DR EMBL: M16CDB101.41;235-240(1995).
DR HSP: D26279; BAA05341.1; -.
DR InterPro: IPR000510; -.
DR Pfam: PF00148; oxidoreductase; 1.
DR NON_TER 136
FT SEQUENCE 136 AA; 15413 MW; E4AA82415AE9CF2A CRC64;

Query Match 48.8%; Score 40; DB 2; Length 136;
Best Local Similarity 46.7%; Pred. No. 5;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Oy 2 VPPLLLEVGVEEF 16

```

Db 115 IKPDLVESGIKEKFI 129
: | | | | | | | | | |

RESULT 18
Q9NW52 PRELIMINARY; PRT; 157 AA.
AC Q9NW52;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE CDNA FLJ10314 FIS, CLONE NT2RM2000374.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi S., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK001176; BAA01534.1; -;
SQ SEQUENCE 157 AA; 16529 MW; 9554BBECCA4300AE CRC64;

Query Match 48.8%; Score 40; DB 4; Length 157;
Best Local Similarity 69.2%; Pred. No. 42;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 VPPLLEVGVEEK 14
: | | | | | | | | | |
Db 88 LPALLREVGVEHK 100

RESULT 19
O28678 PRELIMINARY; PRT; 243 AA.
AC O28678;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE HEME BIOSYNTHESIS PROTEIN (NIRH).
GN AF1594.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton P.W., Spriggs T., Artchak P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL; AE000992; AAB89652.1; -;
DR TIGR; AF1594; -;

KW Hypothetical protein.
SQ SEQUENCE 243 AA; 28252 MW; BC09079DBDD9DC8D CRC64;

Query Match 48.8%; Score 40; DB 1; Length 243;
Best Local Similarity 61.5%; Pred. No. 67;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 4 PLLEVGVEEKFM 16
: | | | | | | | | | |
Db 97 PLLRELGFEEFV 109

RESULT 20
Q9GVK5 PRELIMINARY; PRT; 261 AA.
ID Q9GVK5;
AC Q9GVK5;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HYPOTHETICAL 28.9 KDA PROTEIN (FRAGMENT).
GN LM12.1344.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PRIEDLIN;
RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.,
RA Oliver K.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL390114; CAC02878.1; -;
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 261
SQ SEQUENCE 261 AA; 28909 MW; 8FE2585121982943 CRC64;

Query Match 48.8%; Score 40; DB 5; Length 261;
Best Local Similarity 87.5%; Pred. No. 73;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YVPPLLE 8
: | | | | | | | | | |
Db 130 YVPPLLE 137

RESULT 21
Q9Y8P8 PRELIMINARY; PRT; 348 AA.
ID Q9Y8P8;
AC Q9Y8P8;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HYPOTHETICAL 37.8 KDA PROTEIN APE2585.
GN APE2585.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kwarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
crenarchaeon, Aeropyrum pernix K1.";
RT DNA Res. 6:83-101(1999).

RA La Cognata V., Landschuetz V., Willmitzer L., Mueller-Roeber B.;
 RL Plant pathogen; Phyto. 1990(0).
 CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
 CC OXALOACETATE.
 CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE (BY SIMILARITY).
 CC -1- SIMILARITY: TO THE CITRATE SYNTHASE FAMILY.
 DR EMBL; X84228; CAA59010.1; -.
 DR HSP; P23007; JCSC.
 DR Menden; 152001; 2000; 2977; 15201.
 DR PAM; PF00285; citrate_synth; 1.
 DR PRINTS; PR00143; CITRATSNTASE.
 DR Lysase; Tricarboxylic acid cycle.
 KW NON_TER
 FT
 SQ SEQUENCE 437 AA; 48881 MW; FB50CA21ED5519A3 CRC64;

Query Match 48.8%; Score 40; DB 1; Length 348;
 Best Local Similarity 81.8%; Pred. No. 99;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLLLEVGVEERF 15
 DB 98 LLLRGVEEYF 108

RESULT 22
 Q9N389 PRELIMINARY; PRT; 407 AA.
 AC Q9N389; 2000 (TREMREL. 15, Created)
 DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
 DT 01-OCT-2000 (TREMREL. 15, Last annotation update)
 DE HYPOTHETICAL PROTEIN Y55BIAL.A.
 GN Y55BIAL.A.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 CC Rhabditidae; Eeloderinae; Caenorhabditis.
 RN [1]-taxid=6239;
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE-99069613; PubMed-9851916;
 RA None;
 RT i-Genome sequence of the nematode C. elegans: a platform for
 RL i-Genome biology. The C. elegans Sequencing Consortium. "
 RI Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.; 8-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC024819; AAF59587.1;
 RN [1]-taxid=491;
 RP SEQUENCE 407 AA; 45724 MW; 0B5FF18982C2A848 CRC64;

Query Match 48.8%; Score 40; DB 5; Length 407;
 Best Local Similarity 46.7%; Pred. No. 12e+02;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 LLLRGVEEERF 16
 DB 219 LLAEGLERFI 230

RESULT 23
 Q96544 PRELIMINARY; PRT; 437 AA.
 AC Q96544;
 DT 01-FEB-1997 (TREMREL. 02, Created)
 DT 01-FEB-1997 (TREMREL. 02, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 02, Last annotation update)
 DE CONDENSING ENZYME (EC 4.1.3.7) (CONDENSING ENZYME) (CITRATE
 DE CONDENSING ENZYME) (CITROGENASE) (OXALOACETATE TRANSACETASE)
 DE (FRAGMENT).
 GN CIT1.
 OS Beta vulgaris (Sugar beet).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; Caudicots; Caryophyllidae;
 CC Condensans; Chenopodiaceae; Beta.
 RN [1]
 RP SEQUENCE FROM N.A.

Query Match 48.8%; Score 40; DB 10; Length 437;
 Best Local Similarity 77.8%; Pred. No. 1.3e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 VPLLLEVG 10
 DB 356 VPLLLEVG 364

RESULT 24
 Q9JX05 PRELIMINARY; PRT; 687 AA.
 AC Q9JX05; 2000 (TREMREL. 15, Created)
 DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
 DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
 DE GLYCYL-TRNA SYNTHETASE, BETA CHAIN.
 GN NMB1930.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Eubacteriia; Beta subdivision; Neisseriaceae; Neisseria.
 RN [1]-taxid=491;
 RP SEQUENCE FROM N.A.
 RC STRAIN-MC58 / SEROGROUP B; 10710307;
 RX MEDLINE-20175755; PubMed-10710307;
 RA Tettelin H., Saunders N.J., Hood D.W., Peden J.F., Dodson R.J.,
 Rasmussen J., Klevins M.L., Deboy R., Peterson J.D., Hickey E.K.,
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Mason T., Clecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
 Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamatravala S.L.,
 Gill J., Scarlato V., Masignani E., Rappuoli R., Venter J.C.,
 Smith H.O., Fraser C.A., Nelson K.E., Nelson K.E.,
 RA Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58.
 RL Science 287:1809-1815(2000).
 RN [1]
 RP SEQUENCE 687 AA; 74573 MW; 616BCBDD76A3D4FB CRC64;

Query Match 48.8%; Score 40; DB 2; Length 687;
 Best Local Similarity 48.8%; Pred. No. 2e+02;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 4 PLLLEVGVEERF 16
 DB 259 PVLEAGFEHFL 271

RESULT 25
 Q9JW67


```

DE 464 YVSKLLQEAQIEHKVL 479
RESULT 28
ID Q9K550 PRELIMINARY; PRT: 162 AA.
AC Q9K550;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE X4018 PROTEIN.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
CX NCBI_TaxID=86665;
RN [1] JENCE FROM N.A.
RA Lovell J.; (MAY-2000) to the EMBL/GenBank/DBJ databases.
RX Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RA Takami H., Nakasone K., JCM 9153;
RA STRAIN=C-125 / JCM 9153;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001520; BAB07737.1; -. 119BC6DA25F4118B CRC64;
SQ SEQUENCE 162 AA; 18884 MW; 119BC6DA25F4118B CRC64;

Query Match 47.6%; Score 39; DB 2; Length 162;
Best Local Similarity 54.5%; Pred. No. 65;
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 4 PLLLEVGVEEK 14
DB 151 FVLIELGLEDK 161

RESULT 29
ID Q9U0L8 PRELIMINARY; PRT: 286 AA.
AC Q9U0L8;
DT 01-OCT-2000 (TREMBLrel. 13, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE X4018 PROTEIN.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=5633;
RN [1] JENCE FROM N.A.
RA Devlin K., Bowman S., Churcher C., Harris B., Harris D., Lawson D.,
RA Quail M., Barrell B.; to the EMBL/GenBank/DBJ databases.
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; F01114; CAB62845.2; -.
DE X4018 PROTEIN.
OS Hypothetical protein.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
CX NCBI_TaxID=1219;
RN [1] JENCE FROM N.A.
RA Hess W.R.;
RA MEDLINE=97399394; PubMed-9255521;
RC STRAIN=CCMP 1375;
RC MEDLINE=97399394; PubMed-9255521;
RA "Localization of an open reading frame with homology to human
RA aspartate carboxylase from psba in the prokaryote prochlorococcus
RA DNA Seq. 7:301-306(1997).
RL EMBL; Z80110; CAB02200.1; -.
DR EMBL; Z80110; CAB02200.1; -.
SQ SEQUENCE 286 AA; 33896 MW; 7901PB800BFC485B CRC64;

Query Match 47.6%; Score 39; DB 5; Length 286;
Best Local Similarity 35.7%; Pred. No. 12e+02;
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 YVPLLEVGVEEK 14
DB 76 YIPAIWKYGIER 89

RESULT 30
ID Q95685 PRELIMINARY; PRT: 299 AA.
AC Q95685;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE SERINE-THREONINE SPECIFIC PROTEIN PHOSPHATASE (EC 3.1.3.16)
DE (DJ551D2.3) (PROTEIN PHOSPHATASE 1, REGULATORY SUBUNIT 6).
DE PPPIR6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
CX NCBI_TaxID=9606;
RN [1] JENCE FROM N.A. PubMed-9414128;
RA MEDLINE=98074939; PubMed-9414128;
RX Armstrong C.W., Brown G.J., Cohen P., Cohen P.T.W.;
RA "Identification of a novel member of the family of glycogen-targeting subunits
RA of protein phosphatase 1."
RL FEBS Lett. 418:210-214(1998).
RN [2] JENCE FROM N.A.
RA Lovell J.; (MAY-2000) to the EMBL/GenBank/DBJ databases.
RX Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RA EMBL; AL109528; CAB92096.1; -.
DR EMBL; AL109528; CAB92096.1; -.
SQ SEQUENCE 299 AA; 32559 MW; DB848FB1CF55E49E CRC64;

Query Match 47.6%; Score 39; DB 4; Length 299;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 VPPPLLEVGVEEK 15
DB 242 VPPFLELGSRVHF 255

RESULT 31
ID P72208 PRELIMINARY; PRT: 307 AA.
AC P72208;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE ASPARTOYLASE.
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
CX NCBI_TaxID=1219;
RN [1] JENCE FROM N.A.
RA Hess W.R.;
RA MEDLINE=97399394; PubMed-9255521;
RC STRAIN=CCMP 1375;
RC MEDLINE=97399394; PubMed-9255521;
RA "Localization of an open reading frame with homology to human
RA aspartate carboxylase from psba in the prokaryote prochlorococcus
RA DNA Seq. 7:301-306(1997).
RL EMBL; Z80110; CAB02200.1; -.
DR EMBL; Z80110; CAB02200.1; -.
SQ SEQUENCE 307 AA; 34213 MW; 4711407677490957 CRC64;

Query Match 47.6%; Score 39; DB 2; Length 307;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 5 LILEVGVEEKFM 16
DB 256 LLEGGLEEFV 267

RESULT 32
ID Q9KXR2 PRELIMINARY; PRT: 326 AA.
AC Q9KXR2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)

```

DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)
DE 01-MAR-2001 (TREMELREL. 16, Last annotation update)
DE ASPARTATE CARBAMOYLTRANSFERASE.
GN PYRB
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RT Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL357523; CAB93367.1; -;
DR InterPro: IPR002029; -;
DR PRINTS: PRO0100; AOTCASE.
DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; UNKNOWN_1.
KW Transferase.
SQ SEQUENCE 326 AA; 35661 MW; F353C20B27B43AB0 CRC64;

Query Match 47.6%; Score 39; DB 2; Length 326;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 PPLLEVGVE 12
Db 197 PPTLLPVGVE 206

RESULT 33
QY9727 PRELIMINARY; PRT; 333 AA.
AC Q9P727
DT 01-OCT-2000 (TREMELREL. 15, Created)
DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)
DE 01-MAR-2001 (TREMELREL. 16, Last annotation update)
DE PROBABLE ATP-SPECIFIC SUCCINYL-COA SYNTHETASE ALPHA SUBUNIT.
GN 8D4.130.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO OTHER SUBUNITS (ALPHA) OF SUCCINYL-COA SYNTHETASE,
CC OF MALATE--COA LIGASE AND TO ATP CITRATE-LYASE.
DR EMBL: AL353819; CAB88555.1; -;
DR InterPro: IPR000303; -;
DR Pfam: PF00549; Ligase-CoA; 1.
DR PROSITE: PS01216; SUCCINYL-COA_LIG_1; 1.
DR PROSITE: PS00399; SUCCINYL-COA_LIG_2; 1.

KW Ligase; Phosphorylation.
SQ SEQUENCE 333 AA; 34738 MW; BA6FBF63AD918BDE CRC64;

Query Match 47.6%; Score 39; DB 3; Length 333;
Best Local Similarity 53.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 YVPPLLLEVGVEE 13
Db 106 FVPPPLAAGIEE 118

RESULT 34
P91430 PRELIMINARY; PRT; 419 AA.
ID P91430
AC P91430;
DT 01-MAY-1997 (TREMELREL. 03, Created)
DT 01-MAY-1997 (TREMELREL. 03, Last sequence update)
DT 01-JUN-2000 (TREMELREL. 14, Last annotation update)
DE SIMILAR TO MOLYBDOTERIN BIOSYNTHESIS MOEB PROTEINS.
GN T03FL.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Connell M., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Durbin R., Favello A., Fulton L., Johnston L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Johnston L.,
RA Gardner A., Green P., Hawkins J., Kirsten J., Laister N., Latreille P.,
RA Jones M., Kershaw J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Snaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du Z., Le T.T.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: U88169; ABA4231.1; -;
DR InterPro: IPR00205; -;
DR InterPro: IPR000594; -;
DR InterPro: IPR002162; -;
DR Pfam: PF00899; Thif family; 1.
DR PROSITE: PS00065; D_2_HYDROXYACID_DH_1; UNKNOWN_1.
SQ SEQUENCE 419 AA; 46511 MW; BA61B08DB8AC8F4C CRC64;

Query Match 47.6%; Score 39; DB 5; Length 419;
Best Local Similarity 46.2%; Pred. No. 1.8e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPPLLLEVGVEEK 14
Db 240 VPPPLVVASGIDER 252

RESULT 35
Q21603

[illegible]

```

RESULT 38
Q9H6U0
ID Q9H6U0 PRELIMINARY; PRT; 674 AA.
AC Q9H6U0;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CDNA: FLJ21877 FIS, CLONE HEP02695.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025530; BAB15162.1; -;
SQ SEQUENCE 674 AA; 75965 MW; ELAFD757D82C1AA CRC64;

Query Match 47.6%; Score 39; DB 4; Length 674;
Best Local Similarity 60.0%; Pred. No. 3e+02;
Matches 9; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 VPPLLEVGVEEKFM 16
DB 301 VPPLLCNMG--EKFL 313
||||| :| |||:

RESULT 39
Q9XWL9
ID Q9XWL9 PRELIMINARY; PRT; 800 AA.
AC Q9XWL9;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE Y38F1A.3 PROTEIN.
GN Y38F1A.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wallis J.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken C., Rookpr A., Saunders D., Shownkeen R.,
RA Smailton N., Smith A., Sonhammer E., Staden K., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
DR EMBL; AL032639; CAA21636.1; -;
DR InterPro; IPR000731; -;
DR InterPro; IPR003392; -;
DR Pfam; PF02460; Patched; 1.
SQ SEQUENCE 800 AA; 90816 MW; 738E16F28DE42849 CRC64;

Query Match 47.6%; Score 39; DB 4; Length 1148;
Best Local Similarity 60.0%; Pred. No. 5.2e+02;
Matches 9; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 VPPLLEVGVEEKFM 16
DB 775 VPPLLCNMG--EKFL 787
||||| :| |||:

RESULT 41
Q9LXT9
ID Q9LXT9 PRELIMINARY; PRT; 1963 AA.
AC Q9LXT9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CALLOSE SYNTHASE CATALYTIC SUBUNIT-LIKE PROTEIN.
GN T24H18_170.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Robben J., Grymonprez B., Volckaert G., Bancroft I.,
RA Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL353013; CAB88264.1; -;
DR InterPro; IPR003440; -;
DR Pfam; PF02364; Glucan_synthase; 1.
SQ SEQUENCE 1963 AA; 226393 MW; 242F166954B90C51 CRC64;

Query Match 47.6%; Score 39; DB 10; Length 1963;
Best Local Similarity 50.0%; Pred. No. 9.3e+02;
```

Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
QY 4 PLLLEVGVEERF 15
DB 1628 PMLMEIGLERGF 1639

RESULT 42
Q9LUD7 PRELIMINARY; PRT: 1972 AA.
AC 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE GLUCAN SYNTHASE-LIKE PROTEIN, near cress).
OS Arabidopsis thaliana (strain Col-0).
OC Magnoliophyta: eudicotyledons: Embryophyta: Tracheophyta: Spermatophyta;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN PMLMEIGLERGF 1639
RP SEQUENCE FROM N.A.
RC Sato S, Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLOMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty p1 and-TAC
RT clones.";
RT DNA Res. 7:131-135(2000).
RL EMBL; AB023038; BAB02389.1;
DR InterPro: IPR000051;
KW P22354; Glucan synthase; 1. F30FB9C840A3F43B CRC64;
SQ SEQUENCE 1972 AA; 228012 MW; F30FB9C840A3F43B CRC64;

Query Match 47.6%; Score 39; DB 10; Length 1972;
Best Local Similarity 38.5%; Pred. NO. 9.3e+02; Indels 0; Gaps 0;
Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
QY 4 PLLLEVGVEERF 16
DB 1607 PMVMEIGLERGF 1619

RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004893; AC08192.1;
DR InterPro: IPR000051;
DR PRINTS: PR00196; GEZE: 1.
DR PRINTS: PR00038; HTHLXR.
SQ SEQUENCE 227 AA; 25264 MW; 8043DB0923E969A9 CRC64;

Query Match 46.3%; Score 38; DB 2; Length 227;
Best Local Similarity 46.7%; Pred. NO. 1.4e+02; Indels 0; Gaps 0;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 YVPLLEVGVEERF 15
DB 9 HAPTLLEVERGIAERF 23

RESULT 44
ID 086303 PRELIMINARY; PRT: 259 AA.
AC 086303
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 15, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 15, Last annotation update)
DE MONO-METHYL TRANSFERASE (EC 2.1.1.48).
GN SRMA.
OS Streptomyces ambfaciens.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1889;
RN HAPTLLEVERGIAERF 23
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC23877;
RA Gourmelen A., Blondelet-Rouault M.H., Guerin M., Pernodet J.L.;
RT "Characterization of a macrolide glycosyl transferase gene
RT Streptomyces ambfaciens, the producer of the macrolide antibiotic
RT spiramycin, (FEB-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AJ223970; CAA11706.1;
DR HSSP: P81989; IQE5.
DR InterPro: IPR000051;
DR InterPro: IPR001737;
DR Pfam: PF00398; RnaAD; 1. DIMETH; 1.
DR PROSITE: PS01331; RNAase; 1.
KW RNAase; RNAase; RNAase.
SQ SEQUENCE 259 AA; 28637 MW; B0E61215F73E77E1 CRC64;

Query Match 46.3%; Score 38; DB 2; Length 259;
Best Local Similarity 87.5%; Pred. NO. 1.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 PLLLEVG 10
DB 36 PLLLEVG 43

RESULT 45
Q9ZBW7 PRELIMINARY; PRT: 277 AA.
AC Q9ZBW7
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SC4B3; OLCs coelocolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 46.3%; Score 38; DB 2; Length 259;
Best Local Similarity 87.5%; Pred. NO. 1.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 PLLLEVG 10
DB 36 PLLLEVG 43

Query Match 46.3%; Score 38; DB 2; Length 259;
Best Local Similarity 87.5%; Pred. NO. 1.6e+02; Indels 0; Gaps 0;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 PLLLEVG 10
DB 36 PLLLEVG 43

```

RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MEDLINE-97000351; PubMed-8843436;
RX Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
DR MOL. Microbiol. 21:77-96(1996).
DR EMBL; AL034443; CAA22355.1; -.
DR HSSP; P15121; ZACQ.
DR InterPro; IPR001395; -.
DR Pfam; PF00248; aldo_ket_red; 1.
DR PRINTS; PR00069; ALDKETREDTASE.
DR PROSITE; PS00062; ALDOKETO_REDUCTASE_2; UNKNOWN_1.
DR PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
SQ SEQUENCE 277 AA; 30227 MW; EE7A3D39474B60FE CRC64;

```

```

Query Match          46.3%; Score 38; DB 2; Length 277;
Best Local Similarity 63.6%; Pred. No. 1.7e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 VPLLLEVGVE 12
Db 5 VPPILNNGVE 15

```

Search completed: June 28, 2001, 11:55:30
Job time: 151 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:55:38 ; Search time 72.61 Seconds
(without alignments)
18.221 Million cell updates/sec

Title: US-09-439-313-573

Perfect score: 45

Sequence: 1 MGSIVQLSQS 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database :

1: SPTREMBL16.*

2: sp.archaea.*

3: sp.bacteria.*

4: sp.fungi.*

5: sp.human.*

6: sp.invertebrate.*

7: sp.mammal.*

8: sp.mhc.*

9: sp.organelle.*

10: sp.phage.*

11: sp.plant.*

12: sp.rodent.*

13: sp.unclassified.*

14: sp.virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	75.6	597	Q9HFG9	Q9Hfg9 penicillium
2	33	73.3	3413	2 O54593	O54593 amycolatops
3	32	71.1	408	5 O43978	O43978 babesia bov
4	32	71.1	3469	5 O9U4I2	O9U4I2 drosophila
5	32	71.1	3502	5 O9VYJ9	O9VYJ9 drosophila
6	31	68.9	123	4 Q9Y3N1	Q9Y3N1 homo sapien
7	31	68.9	252	5 Q9VCV0	Q9VCV0 drosophila
8	31	68.9	394	5 O45782	O45782 caenorhabdi
9	31	68.9	436	2 Q9I4F5	Q9I4F5 pseudomonas
10	31	68.9	468	6 O9N135	O9N135 sus scrofa
11	31	68.9	470	6 Q9N129	Q9N129 bos taurus
12	31	68.9	500	2 Q9KWA5	Q9KWA5 agrobacteri
13	31	68.9	617	4 Q9NSZ9	Q9NSZ9 homo sapien
14	30	66.7	68	14 Q9QJG6	Q9QJG6 human immun
15	30	66.7	148	5 Q9NM69	Q9NM69 leishmania
16	30	66.7	161	2 Q9RSZ9	Q9RSZ9 deinococcus
17	30	66.7	170	8 O63327	O63327 bufo verric
18	30	66.7	170	8 O63329	O63329 bufo andrew
19	30	66.7	170	8 O63331	O63331 bufo gargar

ALIGNMENTS

RESULT	1	Q9HFG9	PRELIMINARY;	PRT;	597 AA.
ID	Q9HFG9	AC	Q9HFG9	DT	01-MAR-2001 (TREMREL. 16, Created)
DT	01-MAR-2001	DT	01-MAR-2001	DT	01-MAR-2001 (TREMREL. 16, Last sequence update)
DE	PUTATIVE ALPHA GLUCOSIDASE.	DE	PUTATIVE ALPHA GLUCOSIDASE.	DE	PUTATIVE ALPHA GLUCOSIDASE.
GN	DEXC.	GN	DEXC.	GN	DEXC.
OS	Penicillium minioluteum.	OS	Penicillium minioluteum.	OS	Penicillium minioluteum.
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;	OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;	OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OX	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.	OX	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.	OX	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
RN	[1]	RN	[1]	RN	[1]
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RA	Garcia B., Castellanos A., Menendez J.;	RA	Garcia B., Castellanos A., Menendez J.;	RA	Garcia B., Castellanos A., Menendez J.;
RT	"Isolation of a gene encoding putative alpha glucosidase from	RT	"Isolation of a gene encoding putative alpha glucosidase from	RT	"Isolation of a gene encoding putative alpha glucosidase from
RL	Penicillium minioluteum.";	RL	Penicillium minioluteum.";	RL	Penicillium minioluteum.";
DR	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.	DR	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.	DR	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
SQ	EMBL: AJ278706; CAC09327.1; -. 3F158E544EA70C4 CRC64; SEQUENCE 597 AA; 68120 MW; 3F158E544EA70C4 CRC64;	SQ	EMBL: AJ278706; CAC09327.1; -. 3F158E544EA70C4 CRC64; SEQUENCE 597 AA; 68120 MW; 3F158E544EA70C4 CRC64;	SQ	EMBL: AJ278706; CAC09327.1; -. 3F158E544EA70C4 CRC64; SEQUENCE 597 AA; 68120 MW; 3F158E544EA70C4 CRC64;

20	30	66.7	170	8	O63332	bufo gargar
21	30	66.7	170	8	O63333	bufo gargar
22	30	66.7	170	8	O63957	bufo gargar
23	30	66.7	170	8	O64012	bufo andrew
24	30	66.7	193	8	O21236	reclinomona
25	30	66.7	241	2	Q9HWT0	pseudomonas
26	30	66.7	261	2	Q9KJP3	myxococcus
27	30	66.7	269	14	Q9IP59	tobacco nec
28	30	66.7	277	14	O88611	tobacco nec
29	30	66.7	280	2	Q9PJ60	campylobact
30	30	66.7	327	10	O48641	arabidopsis
31	30	66.7	364	3	Q9P367	neurospora
32	30	66.7	425	10	Q9SV55	arabidopsis
33	30	66.7	473	2	Q9RK75	streptomyce
34	30	66.7	503	2	Q9PDU7	xyella fas
35	30	66.7	539	1	Q9YD39	aeropyrum p
36	30	66.7	587	5	Q9VGP3	drosophila
37	30	66.7	624	5	Q94486	dictyosteli
38	30	66.7	662	4	Q9UFJ0	homo sapien
39	30	66.7	703	4	Q9NQF8	homo sapien
40	30	66.7	750	2	P74111	synecocyst
41	30	66.7	816	5	Q27537	caenorhabdi
42	30	66.7	944	5	Q9VKQ6	drosophila
43	30	66.7	962	10	Q9LJX4	arabidopsis
44	30	66.7	964	10	Q9ZW02	arabidopsis
45	30	66.7	966	10	Q9M6S0	populus tre
46	30	66.7	968	10	Q9ZW07	arabidopsis
47	30	66.7	969	4	Q9Y214	homo sapien
48	30	66.7	972	10	Q9ZW06	arabidopsis
49	30	66.7	1036	10	Q9FPP5	oryza sativ
50	30	66.7	1341	4	Q9UFP8	homo sapien
51	30	66.7	1671	5	Q9NJV5	biomphalari
52	30	66.7	2308	5	Q9VDJ9	drosophila
53	30	66.7	2559	5	O44113	drosophila
54	30	66.7	2559	5	O44381	drosophila
55	30	66.7	2559	5	Q9V7X3	drosophila
56	30	66.7	5060	2	O52545	amycolatops
57	30	66.7	5089	2	O52789	amycolatops
58	29	64.4	115	2	Q9ZFD5	rifta pach
59	29	64.4	125	3	Q9HE79	neurospora
60	29	64.4	126	4	Q9H7W1	homo sapien
61	29	64.4	160	2	Q9K5G3	alcaligenes
62	29	64.4	161	2	Q9K5G2	alcaligenes
63	29	64.4	188	2	Q59802	staphylococ
64	29	64.4	210	2	Q9ZEN6	wolfinella s
65	29	64.4	212	5	Q17223	bombyx mori

Query Match 71.1%; Score 32; DB 5; Length 3469;

Query Match 71.1% Score 32; DB 5; Length 3469;

```
Best Local Similarity 66.7%; Pred. NO. 5.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
Db 291 GGVVQSQS 299

RESULT 5
Q9VYJ9 PRELIMINARY; PRT; 3502 AA.
AC Q9VYJ9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE SMR PROTEIN.
GN SMR OR CG4013.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke S., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fessler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT *The genome sequence of Drosophila melanogaster.*;
RD Science 287:2185-2195(2000).
DR EMBL: AE003490; AAF48196.1;
DR FlyBase: FBgn0024308; Smr.
DR InterPro: IPR000104; -.
DR InterPro: IPR001005; -.
DR InterPro: IPR002086; -.
DR Pfam: PF00249; myb-DNA-binding; 1.
DR PRINTS: PF00308; ANTI-FREZEI
DR PROSITE: PS00687; ALDEHYDE_DEHYDR_GLU; 1.
DR SMART: SM00395; SANT; 1.
```

```
SQ SEQUENCE 3502 AA; 369068 MW; 74C8004F9DA8F8D9 CRC64;

Query Match 71.1%; Score 32; DB 5; Length 3502;
Best Local Similarity 66.7%; Pred. NO. 5.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
Db 422 GGVVQSQS 430

RESULT 6
Q9Y3N1 PRELIMINARY; PRT; 123 AA.
AC Q9Y3N1;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE DJ695020A.1 (PEROXISOME PROLIFERATIVE ACTIVATED RECEPTOR, ALPHA)
DE (FRAGMENT).
GN PPARA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Phillips S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC 1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC 1- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
DR EMBL: AL049856; CAB42862.2; -.
DR HSSP: PI0826; IHRA.
DR InterPro: IPR001628; -.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00047; STROIDFINGER.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
FT NON_TER 123 123
SQ SEQUENCE 123 AA; 12790 MW; 973587F210F505B8 CRC64;

Query Match 68.9%; Score 31; DB 4; Length 123;
Best Local Similarity 60.0%; Pred. NO. 26;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
Db 31 MGNIQISQS 40

RESULT 7
Q9VCV0 PRELIMINARY; PRT; 252 AA.
AC Q9VCV0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CG4803 PROTEIN.
GN CG4803.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA George R.A., Lewis S.E., Ashburner M., Henderson S.N.,
```

RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.Y.,
 RA Watanabe K., Myers E.W., G. Helt G., Nelson C.R., Miklos G.L.G.,
 RA Brail J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Balles R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dev I., Plett S.M.,
 RA Dodson K.J., Evangelista C.C., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Hostin D., Houston K., Howland T.J., W.M.H., Ileguock C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markovits G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M., L.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sinden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J., Wang X.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA Science 287:2195-2195(2000) *Scaphilia melanogaster*.
 CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AE003740; AAF56055.1; -
 DR FlyBase; FBgn0039015; CG4803.
 DR InterPro; IPR000719; -
 DR SMART; SM006290; -
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR SMART; SM00220; S_TKc; 1.
 DR ATP-binding; Serine/threonine-protein kinase; Transferase.
 KW SEQUENCE 252 AA; 28853 MW; 8E006F2EAB58D17 CRC64;
 Query Match 68.9%; Score 31; DB 5; Length 252;
 Best Local Similarity 66.7%; Pred. No. 55;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Oy 2 GSIVOLSQS 10
 Db 155 GTVVDLSQS 163
 RESULT 8
 ID OA5782 PRELIMINARY; PRT; 394 AA.
 AC OA5782;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
 DE T12D8.4 PROTEIN.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RN SEQUENCE FROM N.A.

RA McMurray A.;
 RA Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [12]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=94150718; PubMed=7906398;
 RX Wilson R.J., Alnsough R., Connell M., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Du Z., Durbin R., Favello J., Fulton L.,
 RA Craxton M., Dear S., Ewing R., Hirst J., Kesteven N., Latreille P.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson S., Winkler S., Winkler R.,
 RT 2940; Contiguous nucleotide sequence from chromosome III of C.
 RL Nature 368:32-38(1994).
 DR EMBL; 281120; CAB03344.1; -
 DR EMBL; 394 AA; 44031 MW; 35FE9CEBF8AA028C CRC64;
 SQ SEQUENCE 394 AA; 44031 MW; 35FE9CEBF8AA028C CRC64;
 Query Match 68.9%; Score 31; DB 5; Length 394;
 Best Local Similarity 77.8%; Pred. No. 88;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 MGSIVOLSQ 9
 Db 302 IGSIVOLSQ 310
 RESULT 9
 ID Q914F5 PRELIMINARY; PRT; 436 AA.
 AC Q914F5;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE C4-DICARBOXYLATE TRANSPORT PROTEIN.
 GN DCAF OR PAIL183.
 GN Pseudomonas aeruginosa.
 OS Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OC NCBI_TaxID=287;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=20437337; PubMed=10984043;
 RX Stover C.K., Pham X.-Q.T., Erwin A.L., Miziochichi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.S., Wu Z., Paulsen I.T.,
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen".
 RL Nature 406:959-964(2000).
 DR EMBL; AE004548; AAG04572.1; -
 DR InterPro; IPR001991; -
 DR Pfam; PF00373; SDFTRNSPORT.
 DR PROSITE; PS00714; NA-DICARBOXYL-SYMP-2; 1.
 DR PROSITE; PS00714; NA-DICARBOXYL-SYMP-2; 1.
 SQ SEQUENCE 436 AA; 45993 MW; DCA564DB31A5AC07 CRC64;
 Query Match 68.9%; Score 31; DB 2; Length 436;
 Best Local Similarity 66.7%; Pred. No. 98;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 MGSIVOLSQ 9
 Db 214 VGSIVOLSQ 222

```
RESULT 10
Q9N135
ID Q9N135 PRELIMINARY; PRT; 468 AA.
AC Q9N135;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
DE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR ALPHA (FRAGMENT).
GN PPALPHA.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A.
RA Sundvold H., Frengen A.B., Lien S.;
RT "cdna cloning and expression pattern of the bovine peroxisome
RL proliferator activated receptor (PPAR) alpha and -beta.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
DR EMBL; AF228696; AAF73404.1;
DR InterPro; IPR000536;
DR InterPro; IPR000741;
DR InterPro; IPR001628;
DR InterPro; IPR001723;
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR PRODOM; PD001128; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR SMART; SM00430; HOLI; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
FT NON_TER 468
SQ SEQUENCE 468 AA; 52011 MW; 5812001D8F44BAAA CRC64;

Query Match 68.9%; Score 31; DB 6; Length 468;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
DB 31 MGTQIEISQS 40

RESULT 11
Q9N129
ID Q9N129 PRELIMINARY; PRT; 470 AA.
AC Q9N129;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
DE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR ALPHA (FRAGMENT).
GN PPALPHA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
SEQUENCE FROM N.A.
RA Sundvold H., Frengen A.B., Lien S.;
RT "cdna cloning and expression pattern of the bovine peroxisome
RL proliferator activated receptor (PPAR) alpha and -beta.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: TO C4-TYPE STEROID RECEPTOR ZINC FINGER FAMILY.
DR EMBL; AF229356; AAF73405.1;
DR InterPro; IPR000536;

Query Match 68.9%; Score 31; DB 6; Length 468;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
DB 31 MGTQIEISQS 40
```

```
DR InterPro; IPR001628;
DR InterPro; IPR001723;
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR SMART; SM00430; HOLI; 1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
FT NON_TER 470
SQ SEQUENCE 470 AA; 52436 MW; D205B978EC120A4A CRC64;

Query Match 68.9%; Score 31; DB 6; Length 470;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
DB 31 MGTQIEISQS 40

RESULT 12
Q9KWA5
ID Q9KWA5 PRELIMINARY; PRT; 500 AA.
AC Q9KWA5;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
DE RIORF79 PROTEIN.
GN RIORF79.
OS Agrobacterium rhizogenes.
OG Plasmid pRII724.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=359;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN-MAFF03-01724; PLASMID-PRII724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid,
RT pRII724, by the construction of its physical map and library.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RA STRAIN-MAFF03-01724; PLASMID-PRII724;
RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
RT "Genome structure of Ri plasmid (1). Construction of linking library
RT and physical map of pRII724 in Japanese Agrobacterium.";
RL Nucleic Acids Symp. Ser. 39:189-190(1998).
RN [3]
SEQUENCE FROM N.A.
RA STRAIN-MAFF03-01724; PLASMID-PRII724;
RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
RA Yoshida K.;
RT "The complete nucleotide sequence of a Ri (root inducing) plasmid
RT indicates its chimerical structure between Ti and Sym plasmids.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE FROM N.A.
RA STRAIN-MAFF03-01724; PLASMID-PRII724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid,
RT pRII724, by the construction of its physical map and library.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
SEQUENCE FROM N.A.
RA STRAIN-MAFF03-01724; PLASMID-PRII724;
RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;
RX MEDLINE=20241294; PubMed=10780382;
RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
RA Yoshida K.;
RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and
```

DT 01-MAY-2000 (TREMBLr1.13 Last sequence update)
DT 01-MAY-2001 (TREMBLr1.16 Last annotation update)
DE TRANSSCRIPTION ELONGATION FACTOR.
GN DR1970
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus
OC NCBI_Taxid=1299;
NC 11) SEQUENCE FROM N.A.
NC STRAIN=El RC
RX MEDLINE=20036896; PubMed=10567266;
RX White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen
RA Yamahirova J.J., Lam P., McDonald L., Uitterback T., Zaleswski
RA Makovick K.S., Traving D., Saltschik J., Mikolajczyk, Fenchak
RA Kuchel, K., Nelson K.E., Salzberg S., Smith H.O., Venter J.
RA Fraser C.M.;
RT *Genome sequence of the radioresistant bacterium Deinococcus

```

RT radiodurans RL."
RL Science 286:1571-1577(1999).
DR EMBL: AE002035; AAF11523.1; -.
DR TIGR: DRL970; -.
DR InterPro: IPR001437; -.
DR Pfam: PF01272; GreA_GreB; 1.
DR ProDom: PD004918; -; 1.
KW Elongation factor.
SQ SEQUENCE 161 AA; 18064 MW; 6490C1921015997B CRC64;

Query Match          66.7%; Score 30; DB 2; Length 161;
Best Local Similarity 75.0%; Pred. No. 58;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLS 8
DB 88 LGAIQLS 95

RESULT 17
O63327 PRELIMINARY; PRT; 170 AA.
AC O63327;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo verrucosissimus.
OG Mitochondrion.
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
OX NCBI_TaxID=61429;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SPECIMEN NUMBER: MV2218739;
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tuniyev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal."
RL Mol. Phylogenet. Evol. 9:80-87(1998).
CC -|- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL: AF004526; AAD05131.1; -.
DR InterPro: IPR001750; -.
DR Pfam: PF00361; oxidored_q1; 1.
KW Mitochondrion; NAD; Oxidoreductase.
FT NON_TER 170
SQ SEQUENCE 170 AA; 18220 MW; 9CFD07402DBE6503 CRC64;

Query Match          66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLS 10
DB 139 MALLIQLS 148

RESULT 18
O63329 PRELIMINARY; PRT; 170 AA.
AC O63329;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo andrewsi.

```

```

OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
OX NCBI_TaxID=61428;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SPECIMEN NUMBER: MV2216648;
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tuniyev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal."
RL Mol. Phylogenet. Evol. 9:80-87(1998).
CC -|- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL: AF004529; AAD05137.1; -.
DR InterPro: IPR001750; -.
DR Pfam: PF00361; oxidored_q1; 1.
KW Mitochondrion; NAD; Oxidoreductase.
FT NON_TER 170
SQ SEQUENCE 170 AA; 18225 MW; 2E3705086E28F668 CRC64;

Query Match          66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLS 10
DB 139 MALLIQLS 148

RESULT 19
O63331 PRELIMINARY; PRT; 170 AA.
AC O63331;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo gargarizans.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufo.
OX NCBI_TaxID=30331;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SPECIMEN NUMBER: CAS194318;
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tuniyev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal."
RL Mol. Phylogenet. Evol. 9:80-87(1998).
CC -|- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL: AF004530; AAD05139.1; -.
DR InterPro: IPR001750; -.
DR Pfam: PF00361; oxidored_q1; 1.
KW Mitochondrion; NAD; Oxidoreductase.
FT NON_TER 170
SQ SEQUENCE 170 AA; 18340 MW; 22F619F1E35B5458 CRC64;

Query Match          66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLS 10
DB 139 MALLIQLS 148

```

```

Db 139 MALLIOLSQS 148

RESULT 20
O63332 PRELIMINARY; PRT; 170 AA.
ID O63332
AC O63332
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo gargarizans.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.
NCBI_TaxID=30331;
[1]
SEQUENCE FROM N.A.
MEDLINE=98140324; PubMed=9479697;
Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
Tuniyev B.S., Papenfuss T.J.;
RA Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal. 9:80-87(1998).
RL Mol. phylogenet. Evol. 9:80-87(1998).
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL; AF004531; AAD05141.1; -
InterPro: IPR001750; -
Pfam: PF00361; oxidoreductase.
KW Mitochondrion; NAD; oxidoreductase.
FT NON_TER 170 AA; 18266 MW; 384DBD87BEA7DCB CRC64;
SQ SEQUENCE 170 AA; 18266 MW; 384DBD87BEA7DCB CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
DB 139 MALLIOLSQS 148

RESULT 21
O63333 PRELIMINARY; PRT; 170 AA.
ID O63333
AC O63333
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo gargarizans.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.
NCBI_TaxID=30331;
[1]
SEQUENCE FROM N.A.
MEDLINE=98140324; PubMed=9479697;
Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
Tuniyev B.S., Papenfuss T.J.;
RA Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal. 9:80-87(1998).
RL Mol. phylogenet. Evol. 9:80-87(1998).
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL; AF004533; AAD05145.1; -
InterPro: IPR001750; -
Pfam: PF00361; oxidoreductase.
KW Mitochondrion; NAD; oxidoreductase.
FT NON_TER 170 AA; 18298 MW; 385712C2CA416760 CRC64;
SQ SEQUENCE 170 AA; 18298 MW; 385712C2CA416760 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
DB 139 MALLIOLSQS 148

RESULT 22
O63957 PRELIMINARY; PRT; 170 AA.
ID O63957
AC O63957
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo gargarizans.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.
NCBI_TaxID=30331;
[1]
SEQUENCE FROM N.A.
MEDLINE=98140324; PubMed=9479697;
Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
Tuniyev B.S., Papenfuss T.J.;
RA Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal. 9:80-87(1998).
RL Mol. phylogenet. Evol. 9:80-87(1998).
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL; AF004535; AAD05149.1; -
DR EMBL; AF004532; AAD05143.1; -
DR EMBL; AF004534; AAD05147.1; -
InterPro: IPR001750; -
Pfam: PF00361; oxidoreductase.
KW Mitochondrion; NAD; oxidoreductase.
FT NON_TER 170 AA; 18298 MW; 385712C2CA416760 CRC64;
SQ SEQUENCE 170 AA; 18298 MW; 385712C2CA416760 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
DB 139 MALLIOLSQS 148

RESULT 23
O64012 PRELIMINARY; PRT; 170 AA.
ID O64012
AC O64012
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo andrewsi.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OG Mitochondrion.

```

```

DR InterPro: IPR001750; -
DR Pfam: PF00361; oxidoreductase.
KW Mitochondrion; NAD; oxidoreductase.
FT NON_TER 170 AA; 18284 MW; 38417E1C0589C460 CRC64;
SQ SEQUENCE 170 AA; 18284 MW; 38417E1C0589C460 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
DB 139 MALLIOLSQS 148

RESULT 22
O63957 PRELIMINARY; PRT; 170 AA.
ID O63957
AC O63957
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo gargarizans.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae; Bufo.
NCBI_TaxID=30331;
[1]
SEQUENCE FROM N.A.
MEDLINE=98140324; PubMed=9479697;
Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
Tuniyev B.S., Papenfuss T.J.;
RA Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal. 9:80-87(1998).
RL Mol. phylogenet. Evol. 9:80-87(1998).
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL; AF004535; AAD05149.1; -
DR EMBL; AF004532; AAD05143.1; -
DR EMBL; AF004534; AAD05147.1; -
InterPro: IPR001750; -
Pfam: PF00361; oxidoreductase.
KW Mitochondrion; NAD; oxidoreductase.
FT NON_TER 170 AA; 18298 MW; 385712C2CA416760 CRC64;
SQ SEQUENCE 170 AA; 18298 MW; 385712C2CA416760 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
DB 139 MALLIOLSQS 148

RESULT 23
O64012 PRELIMINARY; PRT; 170 AA.
ID O64012
AC O64012
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT II (FRAGMENT).
GN ND2.
OS Bufo andrewsi.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OG Mitochondrion.

```



```

OC Amphibia; Batrachia; Anura; Neobatrachia; Bufonidae; Bufonidae; Bufo.
OX NCBI_TaxID=61428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPECIMEN NUMBER: CAS194477, SPECIMEN NUMBER: CAS194888;
RX MEDLINE=98140324; PubMed=9479697;
RA Macey J.R., Schulte J.A. II, Larson A., Fang Z., Wang Y.,
RA Tunjiyev B.S., Papenfuss T.J.;
RT "Phylogenetic relationships of toads in the Bufo bufo species group
RT from the eastern escarpment of the Tibetan Plateau: a case of
RT vicariance and dispersal.";
RL Mol. Phylogenet. Evol. 9:80-87(1998).
CC -!- SIMILARITY: TO NADH-UBIQUINONE/PLASTOQUINONE (COMPLEX I), VARIOUS
CC CHAINS.
DR EMBL: AF004528; AD05135.1; -;
DR EMBL: AF004527; AD05133.1; -;
DR InterPro: IPR001750; -;
DR Pfam: PF00361; oxidored_q1; 1.
KW Mitochondrion; NAD; Oxidoreductase.
FT NON_TER 170 170
SQ SEQUENCE 170 AA; 18316 MW; 352DBF02DB58CCD2 CRC64;

Query Match 66.7%; Score 30; DB 8; Length 170;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVLSQS 10
DB 139 MALLIQLS QS 148

RESULT 24
ID Q21236 PRELIMINARY; PRT; 193 AA.
AC Q21236;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE RIBOSOMAL PROTEIN L10.
GN RPL10.
OS Reclinomonas americana.
OG Mitochondrion.
OC Eukaryota; core jakobids; Reclinomonas.
OX NCBI_TaxID=48483;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC50394;
RX MEDLINE=97311393; PubMed=9168110;
RA Lang B.F., Burger G., O'Kelly C.J., Cedergren R., Golding G.B.,
RA Lemieux C., Sankoff D., Turmel M., Gray M.W.;
RT "An ancestral mitochondrial DNA resembling a eubacterial genome in
RT miniature.";
RL Nature 387:493-497(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC50394;
RA Lang B.F., Burger G.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF007261; AAD11863.1; -;
DR InterPro: IPR001790; -;
DR Pfam: PF00466; Ribosomal_L10; 1.
KW Ribosomal protein; Mitochondrion.
SQ SEQUENCE 193 AA; 22448 MW; B1DFCE7D3D7C023F CRC64;

Query Match 66.7%; Score 30; DB 8; Length 193;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 SIVLSQS 10
DB 161 SVVQLTQS 168

```

```

RESULT 25
Q9HWT0
ID Q9HWT0 PRELIMINARY; PRT; 241 AA.
AC Q9HWT0;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE PROBABLE SHORT-CHAIN DEHYDROGENASE.
GN PA4098.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.D., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004826; AAG07485.1; -;
DR InterPro: IPR002198; -;
DR InterPro: IPR002347; -;
DR Pfam: PF00106; adh_short_1;
DR Pfam: PF00678; adh_short_C2; 1.
DR PRINTS; PR00081; GDRDH.
DR PRINTS; PR00080; SDRFAMILY.
DR PROSITE; PS00061; ADH_SHORT; 1.
SQ SEQUENCE 241 AA; 25622 MW; 879AAD48E36F19CE CRC64;

```

```

Query Match 66.7%; Score 30; DB 2; Length 241;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 GSIVLSQS 10
DB 151 GAIVQLTRS 159

RESULT 26
Q9KJP3
ID Q9KJP3 PRELIMINARY; PRT; 261 AA.
AC Q9KJP3;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE 4-OXALOCOTONATE DECARBOXYLASE-LIKE PROTEIN.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=D22;
RX MEDLINE=20032045; PubMed=10564511;
RA Cho K., Zusman D.R.;
RT "Sporulation timing in Myxococcus xanthus is controlled by the espAB
RT locus.";
RL Mol. Microbiol. 34:714-725(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D22;
RA Cho K., Zusman D.R.;
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RL

```

DR EMBL: AF163841; AAF87924.1; --
DR InterPro: IPR002607; Hydratase decarb; 1.
DR Pfam: PF01689; Hydratase decarb; 1.
SQ SEQUENCE 261 AA; 28117 MW; 8562E13FC522FA9A CRC64;

Query Match
Best Local Similarity 66.7%; Score 30; DB 2; Length 261;
Matches 5; Conservative 62.5%; 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSVOLSQ 9
DB 92 GSVOLSQ 99

RESULT 27
ID Q9IP59 PRELIMINARY; PRT; 269 AA.
AC Q9IP59-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE COAT PROTEIN (FRAGMENT).
GN CPT.
OS Tobacco necrosis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OX NCBI_TaxID-12054;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TOYAMA;
RA Sasaki K., Takahashi Y., Oh-oka H., Umeoka T., Oda Y., Fukuyama K.;
RT "The genome sequence of Tobacco necrosis virus (Tobacco necrosis virus strain Toyama) causes of tobacco necrosis disease." Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AB037115; BAB01493.1; --
DR InterPro: IPR000937; --
DR Pfam: PF00729; Viral_coat; 1.
DR InterPro: IPR00233; ICOSAHEDRAL.
DR NON_TER 269
FT NON_TER 269
SQ SEQUENCE 269 AA; 29145 MW; FDA4EF8E26D69C9C CRC64;

Query Match
Best Local Similarity 66.7%; Score 30; DB 14; Length 269;
Matches 7; Conservative 62.5%; 1; Mismatches 0; Gaps 0;

OY 2 GSVOLSQS 10
DB 160 GSVOLSQS 168

RESULT 28
ID Q88611 PRELIMINARY; PRT; 277 AA.
AC Q88611-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-NOV-2000 (Tremblrel. 03, Last annotation update)
DE COAT PROTEIN (FRAGMENT).
OS Tobacco necrosis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OX NCBI_TaxID-12054;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang L.;
RL Thesis (1993), University of Nebraska, Lincoln, USA.
DR EMBL: L04261; CAB23719.1; --
DR HSP: P03607; 4SBV.
DR InterPro: IPR000937; --
DR Pfam: PF00729; Viral_coat; 1.

DR PRINTS: PR00233; ICOSAHEDRAL.
SQ SEQUENCE 277 AA; 30275 MW; 99E81D725B614609 CRC64;

Query Match
Best Local Similarity 65.7%; Score 30; DB 14; Length 277;
Matches 7; Conservative 62.5%; 1; Mismatches 0; Gaps 0;

OY 2 GSVOLSQS 10
DB 168 GSVOLSQS 176

RESULT 29
ID Q9PJ60 PRELIMINARY; PRT; 280 AA.
AC Q9PJ60-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE FLAGELLAR MOTOR SWITCH PROTEIN.
DE FLIY OR CJ0059.
GN Campylobacter jejuni.
OS Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter
OX NCBI_TaxID-197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCTC 11168;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RT "The genome sequence of Campylobacter jejuni strain NCTC 11168." Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL: AB037115; BAB01493.1; --
DR InterPro: IPR001172; --
DR Pfam: PF01052; SPOA; 1.
DR PRINTS: PR00956; FLGMOTORFLIN.
SQ SEQUENCE 280 AA; 31023 MW; 060AC48D485B11B7 CRC64;

Query Match
Best Local Similarity 55.6%; Score 30; DB 2; Length 280;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGSVOLSQ 9
DB 227 IGSVOLSQ 235

RESULT 30
ID Q48641 PRELIMINARY; PRT; 327 AA.
AC Q48641-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DE 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PROTEIN PHOSPHATASE 1 CATALYTIC SUBUNIT.
OS Arabidopsis thaliana (mouse-ear cress).
OC Tracheophyta; Embryophyta; Eudicotyledons; Core eudicots; Rosidae; Arabidopsi
OX NCBI_TaxID-3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Nitsukawa N., Okumura S., Shibata D.;

```

RL Soil Sci. Plant Nutrition 43:971-974(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asanizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones."; 7:31-63(2000).
RL DNA Res. 7:31-63(2000).
CC -!- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +
CC ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC) (BY
CC SIMILARITY).
CC -!- SIMILARITY: TO SERINE/THREONINE SPECIFIC PROTEIN PHOSPHATASE (OR
CC PPP) FAMILY.
DR EMBL; AB000094; BAA24283.1; -.
DR EMBL; AB025638; BAA97417.1; -.
DR HSP; P08129; 1FJM.
DR Mendel; 24637; Arabid;1262;24637.
DR InterPro; IPR000934; -.
DR Pfam; PF00149; Stphosphatase; 1.
DR PRINTS; PR00114; STPHPTASE.
DR PROSITE; PS00125; SER_THR_PHOSPHATASE; UNKNOWN_1.
DR SMART; SM00156; PP2Ac; 1.
KW Hydrolase; Iron; Manganese.
SQ SEQUENCE 327 AA; 37281 MW; CBEEFA363F8C635A CRC64;

Query Match 66.7%; Score 30; DB 10; Length 327;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVLSQS 10
DB 21 GRIVLSET 29

RESULT 31
Q9P367 PRELIMINARY; PRT; 364 AA.
AC Q9P367;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE GR1 RELATED PROTEIN (FRAGMENT).
GN 8D4.260.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Align V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL353819; CAB8568.1; -.
DR InterPro; IPR003592; -.
DR SMART; SM00370; LRR; 1.
FT NON_TER 364 364
SQ SEQUENCE 364 AA; 40305 MW; 069A34B167A461CC CRC64;

Query Match 66.7%; Score 30; DB 3; Length 364;
Best Local Similarity 60.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVLSQS 10
: |:|:|:|:|

```

```

DB 252 IASLVQLAQS 261

RESULT 32
Q9SV55 PRELIMINARY; PRT; 425 AA.
AC Q9SV55;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 44.9 KDA PROTEIN.
GN F25024.30 OR AT4G28910.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078469; CAB43905.1; -.
DR EMBL; AL161573; CAB81479.1; -.
KW Hypothetical protein.
SQ SEQUENCE 425 AA; 44938 MW; 1796FB2FAAB0FB89 CRC64;

Query Match 66.7%; Score 30; DB 10; Length 425;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVLSQS 10
: |:|:|:|:|
DB 279 GGIVALSQS 287

RESULT 33
Q9RK75 PRELIMINARY; PRT; 473 AA.
AC Q9RK75;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PUTATIVE HYDROLASE.
GN SCF11.15.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

```

```

SQ      SEQUENCE      503 AA;  55835 MW;  049EC5CA922F44EB CRC64;
Query Match      Score 30;  PB 2;  Length 503;
Query Local Similarity 66.7%;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps
QY      2 GSVTQLSQS 10
      1 | | | |
Db      417 GSVTQLFQS 425

```

DI 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL_61 kDa PROTEIN APE1073.
DE HYPOTHETICAL_61 kDa PROTEIN APE1073.

INFLUENCE FROM N.A.
 RC SPEARN-KI;
 RX KADLINE=99310339; PubMed-1082966;
 RA kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jinno K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA HosoYama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo H. Y.,
 RA Yamazaki J., Kushida N., Guchai A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako I., Kikuchi H.,
 RA Genom. 2000; 23(1): 101-106. aetabolic hyper-thermophilic
 RT cyerolchase. NERPERYUM bernix Kl.;
 RC CC -1- SIMILARITY: TO DEAD/DEAH BOX HELICASE FAMILY.

```

DR INTERPRO: IPR001410; -
DR INTERPRO: IPR001650; -
DR Pfam: PF00270; DEAD; 1.
DR Pfam: PF00271; helicase_C; 1.
DR PRINTS: PR00851; XRDRMPGMNTB.
DR SMART: SM00490; HELIC_C; 1.
DR ATP-binding: Helicase; Hypothetical protein.
SQ SEQUENCE 539 AA; 61395 MW; 358C78F07CDDDF04 CRC64;

Query Match          66.7%; Score 30; DB 1; Length 539;
Best Local Similarity 70.0%; Pred. NO. 2.1e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps

QY 1 MGSTVQLSOS 10
   M I I I I I
DB 396 MKEIVQLSLE 405

RESULT 36
Q9VGP3 PRELIMINARY; PRT; 587 AA.
ID Q9VGP3
DC 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE CG5723 PROTEIN.
GN CG5723.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Metazoa; Arthropoda; Insecta;
OC Eukaryota; Metazoa; Arthropoda; Insecta;
OC Eukaryota; Metazoa; Arthropoda; Insecta;

```

RC	STRAIN-AX4;
RA	Loomis W.F.;
RL	Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
DR	EMBL; U66913; AAB07544.1; -
FT	NON_TER 624 624
SQ	SEQUENCE 624 AA; 70445 MW; 50343A91ECB0CB37 CRC64;

Query Match	66.7%;	Score 30;	DB 5;	Length 624;
Best Local Similarity	50.0%;	Pred. No. 2.4e+02;		
Matches	5;	Conservative	4;	Mismatches 1; Indels 0; Gaps 0;

Qy	1 MGSIVQLSQS 10
	: : : :
Db	263 LNTLIQLSQS 272

RESULT 38
Q9UFJ0 PRELIMINARY; PRT; 662 AA.
ID Q9UFJ0 AC Q9UFJ0; 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE HYPOTHETICAL 70.3 KDA PROTEIN (FRAGMENT).
GN DKFZP434O225.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Dueterhoeft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117847; CAB56029.1; -
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 662 AA; 70338 MW; D2B3562B55ACFF18 CRC64;

Query Match	66.7%;	Score 30;	DB 4;	Length 662;
Best Local Similarity	85.7%;	Pred. No. 2.6e+02;		
Matches	6;	Conservative	1;	Mismatches 0; Indels 0; Gaps 0;

Qy	2 GSIVOLS 8
	:::
Db	590 GSVVOLS 596

RESULT 39
Q9NQF8 PRELIMINARY; PRT; 703 AA.
ID Q9NQF8 AC Q9NQF8; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE DJ1184F4.1 (KIAA0978) (FRAGMENT).
GN DJ1184F4.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL034550; CAC00607.1; -
FT NON_TER 1 1
SQ SEQUENCE 703 AA; 74741 MW; 8EE4B3F35672F012 CRC64;

Query Match	66.7%;	Score 30;	DB 4;	Length 703;
-------------	--------	-----------	-------	-------------

Best Local Similarity 85.7%; Pred. No. 2.7e+02; Indels 0; Gaps 0; Mismatches 1; Conservative 6; Matches 6;

QY 2 GSVOLQS 8
DB 631 GSVOLQS 637

RESULT 40

AC P74111 PRELIMINARY; PRT: 750 AA.
ID 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE SENSORY TRANSDUCTION HISTIDINE KINASE.
GN SLR1969.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RX NCBI_TaxID=1148;
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Suglura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Tabata S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC KINASES.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES
DR EMBL: D90912; BAA18197.1;
DR InterPro: IPR000410;
DR InterPro: IPR001294;
DR InterPro: IPR001789;
DR InterPro: IPR002018;
DR Pfam: PF00512; signal; 1.
DR Pfam: PF01590; GAF; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
DR SMART: SM00448; REC; 1.
KW Kinase; Phosphorylation; Sensory transduction; Transferase.
SQ SEQUENCE 750 AA; 83989 MW; DJDCB538F421896A CRC64;

Query Match 66.7%; Score 30; DB 2; Length 750;
Best Local Similarity 66.7%; Pred. No. 2.9e+02; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQ 9
DB 437 INSIILSQ 445

RESULT 41

AC Q27537 PRELIMINARY; PRT: 816 AA.
ID 027537;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE SENSORY TRANSDUCTION HISTIDINE KINASE.
GN SLR1969.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RX NCBI_TaxID=1148;
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Suglura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Tabata S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC KINASES.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
CC KINASES
DR EMBL: D90912; BAA18197.1;
DR InterPro: IPR000410;
DR InterPro: IPR001294;
DR InterPro: IPR001789;
DR InterPro: IPR002018;
DR Pfam: PF00512; signal; 1.
DR Pfam: PF01590; GAF; 1.
DR PROSITE: PS50046; PHYTOCHROME_2; 1.
DR SMART: SM00448; REC; 1.
KW Kinase; Phosphorylation; Sensory transduction; Transferase.
SQ SEQUENCE 750 AA; 83989 MW; DJDCB538F421896A CRC64;

RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA MEDLINE=94150718; PubMed=7906398;
RA Wilton R., Alnsough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Craxton M., Dear S., Durbin H., Hillier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons N., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Shalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierri-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilson R., Young I., Zannis-Hartmann S.,
RT "240bp of contiguous nucleotide sequence from chromosome III of C.
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS (BY SIMILARITY).
CC CATALYTIC SUBUNIT. DNTP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS (BY SIMILARITY).
CC -1- SIMILARITY: TO DNA TOPOISOMERASE II FAMILY.
DR EMBL: D90912; BAA18197.1;
DR HSP; P06786; IPR001241;
DR InterPro: IPR002205;
DR InterPro: IPR002205;
DR Pfam: PF00204; DNA_topoisomII; 1.
DR Pfam: PF00521; DNA_topoisomII; 1.
DR PRODOM: PD000616; TOPOISOMERASE_II; 1.
DR SMART: SM00434; TOP4C; 1.
KW ATP-binding; DNA-binding; Isomerase; Topoisomerase.
SQ SEQUENCE 816 AA; 93029 MW; B3A0A6C61D061AEB CRC64;

Query Match 66.7%; Score 30; DB 5; Length 816;
Best Local Similarity 66.7%; Pred. No. 3.2e+02; Indels 0; Gaps 0;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQ 9
DB 373 MCTIVNLAQ 381

RESULT 42

Q9VK06 PRELIMINARY; PRT: 944 AA.
ID Q9VK06;
AC Q9VK06;
DT 01-MAY-2000 (TREMBlrel. 13, Created)

01-MAY-2000 (Tremblrel. 13, Last sequence update)
01-MAY-2000 (Tremblrel. 13, Last annotation update)
CG6729 PROTEIN.
CG6729.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003629; AAF53005.1; -
DR FlyBase; FBgn0032296; CG6729.
SQ SEQUENCE 944 AA; 106827 MW; E02D6405B35645BF CRC64;

Query Match 66.7%; Score 30; DB 5; Length 944;
Best Local Similarity 75.0%; Pred. No. 3.7e+02;
Matches 9; Conservative 1; Mismatches 0; Indels 2; Gaps 1;
QY 1 MGSIV-QLSQS 10
| | | | | : | | | |
Db 865 MGSIVPRLS QS 876

RESULT 43
Q9LJX4 PRELIMINARY; PRT; 962 AA.
AC Q9LJX4
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE RNA BINDING PROTEIN-LIKE.

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL; AP000383; BAB01884.1; -
DR InterPro; IPR001313; -
DR Pfam; PF00806; PUF; 8;
SQ SEQUENCE 962 AA; 107201 MW; E95B8CBEB375D2F2 CRC64;

Query Match 66.7%; Score 30; DB 10; Length 962;
Best Local Similarity 62.5%; Pred. No. 3.8e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 2 GSIVQLSQ 9
| | | | | : | | | |
Db 849 GNVVQMSQ 856

RESULT 44
Q9ZW02 PRELIMINARY; PRT; 964 AA.
AC Q9ZW02;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUMILIO-LIKE PROTEIN.
GN FL6P2.48
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLOMBIA;
RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
RA Shen M., Renning C.M., Fraser C.M., Somerville C.R., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC F16P2 genomic sequence."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004561; AAC95220.1; -
DR InterPro; IPR001313; -
DR Pfam; PF00806; PUF; 8;
DR SMART; SM00025; Pumilio; 1.
SQ SEQUENCE 964 AA; 105986 MW; 40E3FFB1B5A6B43A CRC64;

Query Match 66.7%; Score 30; DB 10; Length 964;
Best Local Similarity 75.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 GSIVQLSQ 9
| | | | | : | | | |
Db 846 GNVVQMSQ 853

RESULT 45
Q9M6S0

ID Q9M6SO PRELIMINARY; PRT; 966 AA.
 AC Q9M6SO;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE PUMILIO DOMAIN-CONTAINING PROTEIN PPD1.
 GN PPD1.
 OS Populus tremula x Populus tremuloides.
 OC Eukaryota; Viridiplantae; Eudicotyledons; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurossids I;
 OC Malpighiales; Salicaceae; Populus.
 CX NCBI_TaxID=47664;
 OX [1]
 RN SEQUENCE FROM N.A.
 RA "A pumilio RNA binding domain-containing protein is expressed in
 RT vascular tissue of Hybrid Aspen";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF153276; AAF71823.1;
 DR InterPro; IPR001313;
 DR RefSeq; NP_060001.2;
 DR PDB; 1P0806; POP2; 8;
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; UNKNOWN 1.
 SQ SEQUENCE 966 AA; 105324 MW; 728CF2320C28E652 CRC64;

Query Match 56.78; Score 30; DB 10; Length 966;
 Identical 100%; 1; Mismatches 1; Indels 0; Gaps 0;
 Matches 6; Conservative

Qy 2 GSIVQLSQ 9
 Db 884 GKIVMSQ 891

Search completed: June 28, 2001, 11:55:40
 Job time: 161 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:54:15 ; Search time 66.55 Seconds
(without alignments)
24.596 Million cell updates/sec

Title: US-09-439-313-566
Perfect score: 146
Sequence: 1 VGGGLYQGVPRAPTEGARRHYDEGYR 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : A_Geneseq_0601.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146	100.0	553	AAW71869	Amino acid encoded
2	146	100.0	553	AAW69385	Prostate tumour sp
3	146	100.0	553	AAW28527	Protein encoded by
4	146	100.0	553	AAW82002	Human immunogenic
5	131	89.7	27	AAW85071	PSI08 protein deri
6	131	89.7	27	AAW29271	Human prostate-rel
7	131	89.7	255	AAW85068	Protein encoded by
8	131	89.7	255	AAW29268	Human prostate-rel
9	63	43.2	742	AAW94977	Human secreted pro
10	50	34.2	27	AAW09155	Hepatitis GB virus
11	50	34.2	3164	AAW94346	Hepatitis GB virus

RESULT 1	AAW71869	AAW71869 standard; Protein; 553 AA.
ID	AAW71869	
XX	XX	AAW71869;
AC	AC	AAW71869;
XX	XX	AAW71869;
DT	DT	06-JAN-1999 (first entry)
XX	XX	AAW71869;
DE	DE	Amino acid encoded by prostate tumour clone L1-12.
XX	XX	Prostate; cancer; tumour; vaccine; immunogen; clone.
KW	KW	Prostate; cancer; tumour; vaccine; immunogen; clone.
OS	OS	Homo sapiens.
XX	XX	AAW71869;

ALIGNMENTS

ze mays protein f
ze mays protein f
Wheat sucrose tran
Amino acids 145-23
Amino acid sequenc
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
The amino acid seq
Secreted protein C
Secreted protein C
A human FAN (facto
Phosphoribosylform
Monkey p53 tumour
Laccase RSLac3. R
Rhizoctonia solani
Rhizoctonia solani
Rhizoctonia solani
R. solani laccase
Eucalyptus grandis
Human multiple end
Tumor suppressor p
Human OREF318
Monkey p53 tumour
Plasmid pASK75 ope
Plasmid pASK60-Str
Rat Fas receptor.
Pseudomonas sp. HR
Pseudomonas sp. HR
Human oxidoreducta
Arabidopsis thalia
Arabidopsis thalia
Human cancer assoc
A. terreus LovC de
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Alpha 2-Macroglobu
Human alpha-2-MR.
Mutant full-length
Human complement r
Human decay accele
Decay accelerating
Decay accelerating
Human decay accele
Human CD55 and 791

PN WO9837093-A2.
 XX 27-AUG-1998.
 XX 25-FEB-1998; 98WO-US03492.
 XX 09-FEB-1999; 98US-0020956.
 XX 25-FEB-1999; 97US-0806099.
 XX 01-AUG-1997; 97US-0904804.
 XX (CORI-) CORIXA CORP.
 XX Dillon DC, Xu J;
 XX WPI: 1998-609886/51.
 XX N-PSDB; AAV61201.
 XX Polypeptides comprising immunogenic portions of prostate proteins -
 PT used in a vaccine for the treatment of prostate cancer
 XX
 XX Example 1; Page 82-84; 130pp; English.
 XX The present sequence is an immunogenic portion of a prostate tumour
 CC protein. The immunogen, or the DNA encoding it, can be used as a
 CC vaccine for the treatment of prostate cancer. The immunogen was
 CC isolated from a prostate tumour cDNA library obtained by subtracting
 CC a prostate tumour cDNA expression library with a normal tissue cDNA
 CC library.
 XX
 XX SQ Sequence 553 AA;
 Query Match 100.0%; Score 146; DB 19; Length 553;
 Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Gaps 0;
 Matches 27; Conservative 0; Mismatches 0;
 QY 1 VGEGLYQGVPRAEPTGTEARRHYDEGVR 27
 Db 296 vgeglyqgvpraeptgtearrhydegvr 322
 ||||||||||||||||||||||||||||
 RESULT 2
 ID AAW69385 standard; Protein: 553 AA.
 XX AAW69385;
 XX 08-DEC-1998 (first entry)
 XX Prostate tumour specific gene clone L1-12 protein.
 XX Prostate tumour specific gene; human; prostate cancer; detection;
 XX therapy.
 XX Homo sapiens.
 XX WO9837418-A2.
 XX 27-AUG-1998.
 XX 25-FEB-1998; 98WO-US03690.
 XX 09-FEB-1998; 98US-0904809.
 XX 25-FEB-1997; 97US-0806596.
 XX 01-AUG-1997; 97US-0904809.
 XX (CORI-) CORIXA CORP.
 XX Dillon DC, Xu J;
 XX WPI: 1998-480805/41.
 XX N-PSDB; AAV58586.

PT Novel human prostate specific tumour protein and fragments - useful
 PT for detecting and treating prostate cancers
 XX
 XX Example 1; Page 87-89; 141pp; English.
 XX This sequence is encoded by a human prostate tumour specific gene, and
 CC can be used in the method of the invention. The method is for detecting
 CC prostate cancer comprises contacting a biological sample with an agent
 CC able to bind an immunogenic portion of a prostate protein (such as
 CC this protein sequence). An antibody which binds to an immunogenic
 CC portion of the prostate protein, and the method can be used to detect,
 CC monitor progression of, or treat prostate cancers. The antibody may
 CC also be conjugated to a therapeutic agent for use in therapy of prostate
 CC cancers.
 XX
 XX SQ Sequence 553 AA;
 Query Match 100.0%; Score 146; DB 19; Length 553;
 Best Local Similarity 100.0%; Pred. No. 0; Indels 0; Gaps 0;
 Matches 27; Conservative 0; Mismatches 0;
 QY 1 VGEGLYQGVPRAEPTGTEARRHYDEGVR 27
 Db 296 vgeglyqgvpraeptgtearrhydegvr 322
 ||||||||||||||||||||||||||||
 RESULT 3
 ID AAB28527 standard; Protein: 553 AA.
 XX AAB28527;
 XX 07-FEB-2001 (first entry)
 XX Protein encoded by human breast tumour cDNA clone P501S.
 XX Human; breast tumour antigen; cytostatic; immunotherapy;
 XX breast cancer; vaccine.
 XX Homo sapiens.
 XX WO200061756-A2.
 XX 19-OCT-2000.
 XX 10-APR-2000; 2000WO-US09688.
 XX 09-APR-1999; 99US-0288950.
 XX 02-JUL-1999; 99US-0346327.
 XX (CORI-) CORIXA CORP.
 XX Reed SG, Xu J, Dillon DC;
 XX WPI: 2000-638568/61.
 XX N-PSDB; AAC79473.
 XX A novel isolated polypeptide comprising an immunogenic portion of a
 PT breast cancer protein useful in the detection and treatment of breast
 PT cancer -
 XX
 XX Claim 2; Page 92-93; 95pp; English.
 XX The present sequence is encoded by a cDNA sequence which was isolated
 CC from a breast tumour cDNA library. It is provided in a specification
 CC relating to compounds for immunotherapy and diagnosis of breast cancer.
 CC Breast tumour antigens and polypeptides containing immunogenic
 CC epitopes of a breast cancer protein are provided. A pharmaceutical composition to be used in the
 CC treatment of breast cancer. Proliferated T cells and incubated antigen
 CC presenting cells are also required. The polypeptides and polynucleotides
 CC may also be used to produce a vaccine.

```
SQ Sequence 553 AA;
      Query Match      100.0%; Score 146; DB 21; Length 553;
      Best Local Similarity 100.0%; Pred. No. 4.8e-13;
      Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGEGLYQGVPRAPGTEARRHYDEGVR 27
    |||||
Db 296 vgeglyqgvpraepgtearrhydegvr 322

RESULT 4
AAY82002
ID AAY82002 standard; Protein; 553 AA.
XX
AC AAY82002;
XX
DT 13-JUN-2000 (first entry)
XX
DE Human immunogenic prostate tumour protein sequence SEQ ID NO:113.
XX
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW immunogenic; cytostatic; vaccine.
XX
OS Homo sapiens.
XX
PN WO200004149-A2.
XX
PD 27-JAN-2000.
XX
PF 14-JUL-1999; 99WO-US15838.
XX
PR 14-JUL-1998; 98US-0115453.
PR 14-JUL-1998; 98US-0116134.
PR 23-SEP-1998; 98US-0159812.
PR 23-SEP-1998; 98US-0159822.
PR 15-JAN-1999; 99US-0232149.
PR 15-JAN-1999; 99US-0232880.
PR 09-APR-1999; 99US-0288946.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
XX
DR WPI; 2000-171268/15.
XX
PT New polypeptide useful for treating and diagnosing prostate cancer
PT comprises an immunogenic portion of prostate tumor protein -
XX
PS Claim 3; Page 138-139; 263pp; English.
XX
CC The present invention describes isolated polypeptides, comprising an
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC and polynucleotides encoding them have cytostatic activity and can be
CC used in vaccines and in gene therapy. The polypeptides and
CC polynucleotides encoding them, antigen presenting cells which express
CC the polypeptides, antibodies against the polypeptides and vaccines
CC comprising them can be used for inhibiting the development of prostate
CC cancer in a patient. The polypeptides can be used to generate antibodies
CC or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC the polynucleotides encoding the polypeptides can be used as a probe or
CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC AAY82000 to AAY82020 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 553 AA;

      Query Match      100.0%; Score 146; DB 21; Length 553;
      Best Local Similarity 100.0%; Pred. No. 4.8e-13;
      Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGEGLYQGVPRAPGTEARRHYDEGVR 27
    |||||
Db 296 vgeglyqgvpraepgtearrhydegvr 322

RESULT 5
AAW85071
ID AAW85071 standard; Peptide; 27 AA.
XX
AC AAW85071;
XX
DT 12-FEB-1999 (first entry)
XX
DE PS108 protein derived peptide.
XX
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW drug screening; gene therapy.
XX
OS Homo sapiens.
XX
PN WO9850567-A1.
XX
PD 12-NOV-1998.
XX
PF 01-MAY-1998; 98WO-US08930.
XX
PR 02-MAY-1997; 97US-0850713.
XX
PA (ABBO ) ABBOTT LAB.
XX
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;
PI Roberts-Rapp L, Russell JC, Stroupe SD;
XX
DR WPI; 1999-034731/03.
XX
PT New isolated prostate-specific polynucleotides - used to develop
PT products for the diagnosis and treatment of prostate diseases, e.g.
PT benign hyperplasia, prostatic or prostate cancer
XX
PS Claim 17; Page 101; 122pp; English.
XX
CC The present sequence is derived from the PS108 protein. The
CC specification describes a method for detecting the presence
CC of a target PS108 polynucleotide in a test sample. The method
CC comprises contacting the test sample with at least 1 PS108-specific
CC polynucleotide or complement, and detecting the presence of the target
CC PS108 polynucleotide. The products can be used for detecting, diagnosing,
CC staging, monitoring, prognosticating, in vivo imaging, preventing or
CC treating, or determining predisposition to diseases or conditions of
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC products can be used in drug screening and gene therapy.
XX
SQ Sequence 27 AA;

      Query Match      89.7%; Score 131; DB 20; Length 27;
      Best Local Similarity 100.0%; Pred. No. 2.6e-12;
      Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GLYQGVPRAPGTEARRHYDEGVR 27
    |||||
Db 1 glyqgvpraepgtearrhydegvr 24

RESULT 6
AAB29271
ID AAB29271 standard; Protein; 27 AA.
XX
AC AAB29271;
XX
```

DT 07-FEB-2001 (first entry)
 DE Human prostate-related PS108 partial protein sequence SEQ ID NO: 39.
 KW Human; prostate cancer; PS108; antibody; tumour; metastasis.
 XX Homo sapiens.
 XX US6130043-A.
 XX 10-OCT-2000.
 PF 01-MAY-1998; 98US-0071710.
 PR 02-MAY-1997; 97US-0850713.
 XX (ABBO) ABBOTT LAB.
 PA Cohen M, Colpitts TL, Friedman PN, Gordon J, Granados EN;
 PI Billing-Medel PA, Klass MR, Roberts-Rapp L, Stroupe SD, Yu H;
 KR Kratochvil JD, Russell JC, Hodges SC;
 DR WPI: 2000-655655/63.
 XX Methods for detecting target prostate-specific polynucleotides or
 PT diseases of the prostate (e.g. prostate cancer), comprising detecting
 XX the presence of any of PS108 nucleic acid sequences in a test sample -
 XX Example 10; Column 87-88; 55pp; English.
 PS The present invention is related to a number of partial coding and
 CC protein sequences for the human prostate tissue protein PS108. These
 CC sequences can be used in the diagnosis and prognosis of prostate
 CC diseases. Particularly, prostate cancer can be used to produce
 CC antisera, which can be used in the treatment. The present sequence is one
 CC of the PS108 partial protein sequences.
 XX Sequence 27 AA;
 SQ
 Query Match 89.7%; Score 131; DB 21; Length 27;
 Best Local Similarity 100.0%; Pred. No. 2, 6e-12; Indels 0; Gaps 0;
 Matches 24; Conservative 0; Mismatches 0;
 QY 4 GLYQGVPRAPGTEARRHYDEGVR 27
 Db 1 glyqgvprapgtearrhydegvr 24
 RESULT 7
 AA085068
 ID AAW85068 standard; Protein: 255 AA.
 XX
 XX AAW85068;
 XX 12-FEB-1999 (first entry)
 DE Protein encoded by the consensus sequence of the PS108 gene.
 XX PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
 KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
 KR drug screening; gene therapy.
 XX Homo sapiens.
 OS WO9805067-A1.
 XX 12-NOV-1998.
 PD 01-MAY-1998; 98WO-US08930.
 PR 02-MAY-1997; 97US-0850713.
 XX

PA (ABBO) ABBOTT LAB.
 XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
 PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;
 KR Roberts-Rapp L, Russell JC, Stroupe SD;
 DR WPI: 1999-034731/03.
 XX N-PSDB; AAV71181.
 XX New isolated prostate-specific polynucleotides - used to develop
 PT products for the diagnosis and treatment of prostate diseases, e.g.
 PT benign hyperplasia, prostatic or prostate cancer
 XX Claim 17; Pages 99-100; 122pp; English.
 CC The present sequence is encoded by the consensus sequence for a PS108
 CC gene. The sequence is derived from overlapping clones AAV71166-79. The
 CC clone sequences are PS108 gene-specific. They are used in the method
 CC of the invention. The specification describes a method for detecting the
 CC presence of the PS108 gene in a test sample. The method
 CC involves contacting the test sample with at least 1 PS108-specific
 CC polynucleotide or complement, and detecting the presence of the target
 CC polynucleotide. The products can be used for detecting,
 CC diagnosing, staging, monitoring, prognosticating, in vivo imaging,
 CC preventing or treating, or determining predisposition to diseases or
 CC conditions of the prostate such as hyperplasia (BPH),
 CC prostatitis, prostatic intraepithelial neoplasia (PIN) and cancer in
 CC particular. The products can be used in drug screening and gene therapy.
 XX Sequence 255 AA;
 SQ
 Query Match 89.7%; Score 131; DB 20; Length 255;
 Best Local Similarity 100.0%; Pred. No. 3e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 GLYQGVPRAPGTEARRHYDEGVR 27
 Db 1 glyqgvprapgtearrhydegvr 24
 RESULT 8
 AAB29268
 ID AAB29268 standard; Protein: 255 AA.
 XX
 XX AAB29268;
 XX 07-FEB-2001 (first entry)
 DE Human prostate-related PS108 partial protein sequence SEQ ID NO: 36.
 KW Human; prostate cancer; PS108; antibody; tumour; metastasis.
 XX Homo sapiens.
 OS US6130043-A.
 XX 10-OCT-2000.
 PF 01-MAY-1998; 98US-0071710.
 PR 02-MAY-1997; 97US-0850713.
 XX (ABBO) ABBOTT LAB.
 PA Cohen M, Colpitts TL, Friedman PN, Gordon J, Granados EN;
 PI Billing-Medel PA, Klass MR, Roberts-Rapp L, Stroupe SD, Yu H;
 KR Kratochvil JD, Russell JC, Hodges SC;
 DR WPI: 2000-655655/63.
 XX Methods for detecting target prostate-specific polynucleotides or
 PT diseases of the prostate (e.g. prostate cancer), comprising detecting

PT the presence of any of PS108 nucleic acid sequences in a test sample -
 XX
 PS Example 1; Column 85-88; 53pp; English.
 XX
 CC The present invention is related to a number of partial coding and
 CC protein sequences for the human prostate tissue protein PS108. These
 CC sequences can be used in the diagnosis and prognosis of prostate
 CC diseases, particularly prostate cancer. They can also be used to produce
 CC antibodies which can be used in treatment. The present sequence is one
 CC of the PS108 partial protein sequences.
 XX
 SQ Sequence 255 AA;
 Query Match 89.7%; Score 131; DB 21; Length 255;
 Best Local Similarity 100.0%; Pred. No. 3e-11; Mismatches 0; Indels 0; Gaps 0;
 Matches 24; Conservative 0;
 QY 4 GLYQGVPRAPGTEARRHYDEGVR 27
 DB 1 glyqgvpraepgtearrhydegvr 24
 RESULT 9
 AAY94977
 ID AAY94977 standard; Protein; 742 AA.
 XX
 AC AAY94977;
 XX
 DT 16-JUN-2000 (first entry)
 XX
 DE Human secreted protein clone as180_1 protein sequence SEQ ID NO:160.
 XX
 KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiaschmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy.
 XX
 OS Homo sapiens.
 XX
 PN WO200009552-A1.
 XX
 PD 24-FEB-2000.
 XX
 PF 13-AUG-1999; 93WO-US18298.
 XX
 XX 14-AUG-1998; 98US-0096622.
 PR 17-AUG-1998; 98US-0096815.
 PR 04-SEP-1998; 98US-0099229.
 PR 23-OCT-1998; 98US-0105368.
 PR 08-JAN-1999; 99US-0115234.
 PR 12-FEB-1999; 99US-0119931.
 PR 18-FEB-1999; 99US-0120575.
 PR 30-APR-1999; 99US-0132020.
 PR 11-AUG-1999; 99US-0096622.
 XX
 XX (GEMY) GENETICS INST INC.
 XX
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark HF, Fechtel K;
 XX
 DR WPI; 2000-205979/18.
 XX
 XX New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin

PT antiinflammatory or tumor inhibition activity -
 XX
 PS Claim 169; Page 614-616; 641pp; English.
 XX
 CC AAA16618 to AAA16697 encode the human secreted proteins given in
 CC AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
 CC probes for the human secreted proteins from the present invention.
 XX
 SQ Sequence 742 AA;
 Query Match 43.2%; Score 63; DB 21; Length 742;
 Best Local Similarity 51.7%; Pred. No. 0.66;
 Matches 15; Conservative 2; Mismatches 8; Indels 4; Gaps 1;
 QY 3 EGYQGVPRAPGTEARRHYD----EGVR 27
 DB 95 egeyggipraesgkggermadgaplagvr 123
 RESULT 10
 AAB09155
 ID AAB09155 standard; Protein; 27 AA.
 XX
 AC AAB09155;
 XX
 DT 30-AUG-2000 (first entry)
 XX
 DE Hepatitis GB virus protein sequence SEQ ID NO:277.
 XX
 KW Hepatitis GB virus; HGBV; diagnosis; therapeutic; immunogenic;
 KW infection; detection; characterisation; hepatitis.
 XX
 OS Hepatitis GB virus.
 XX
 PN US6051374-A.
 XX
 PD 18-APR-2000.
 XX
 PF 07-JUN-1995; 95US-0488445.
 XX
 XX 14-FEB-1994; 94US-0196030.
 PR 13-MAY-1994; 94US-0242654.
 PR 29-JUL-1994; 94US-0283314.
 PR 23-NOV-1994; 94US-0344185.
 PR 23-NOV-1994; 94US-0344190.
 PR 30-JAN-1995; 95US-0377557.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Dawson GJ, Leary TP, Muerhoff AS, Pilot-Matias TJ, Buijk SL;
 PI Mushahwar IK, Simons JN, Desai SM, Erker JC, Schlauder GG;

Muerhoff AS, Mushahwar IK, Pilot-Matias TJ, Schlauder GG;
Simons JN;

WPI: 1995-293123/38.
N-PSDB; AAT00129.

Non-A, non-B, non-C, non-D, non-E Hepatitis virus reagents - useful
for diagnosis and therapy of hepatitis GB virus

Example 9; Pages 343-357; 661pp; English.

Double tetrads hepatitis B virus (HBV) DNA obtd. from HGBV
infected tamarin plasma, using standard procedures, was used to
prepare a lambda phage HGBV cDNA library. Clones were rescued
from the lambda phage, searched against a sequence database and
sequenced. The unique open reading frame was identified. This
sequence was compared with the sequences AAT00129/30 (GB contig A and B) which encode
the proteins AAR94345-47 (the 3 possible coding strand reading
frames) and AAR82072, respectively. Reagents which comprise the HGBV
DNA, or its protein prods. can be used for the diagnosis, therapy
or in a vaccine to prevent HGBV infection.

Sequence 3164 AA:

Query Match 34.2%; Score 50; DB 16; Length 3164;
Best Local Similarity 43.5%; Pred. NO. 2.4e+02;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

OY 4 GLYGVPRAEPGTEARHYDEGV 26
|||
|||
308 glgdyphqspcltcsqhlregl 330

RESULT 12
AAG32765
ID AAG32765 standard; Protein; 113 AA.
XX AAG32765;
XX
DT 17-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 39589.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
XP EP1033405-A2.
PN
PD
PP 06-SEP-2000.
PX
PY 25-FEB-2000; 2000EP-0301439.
PR
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126284.
PR 01-APR-1999; 99US-0127264.
PR 01-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130801.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

[illegible]


```

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 32.9%; Score 48; DB 21; Length 123;
Best Local Similarity 54.5%; Pred. No. 14;
Matches 12; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

QY 3 EGLY----QGVPRAPGTEARR 20
    ||| : ||||| :|||
Db 53 egllrrdravaraepgrqarr 74

RESULT 14
AA41125
ID AA41125 standard; protein; 563 AA.
XX
AC AA41125;
XX
DT 17-JAN-2000 (first entry)
XX
DE Wheat sucrose transport protein (clone wlmkl.pk0002.e11).
XX
KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;
KW carbohydrate transport; plant cell.
XX
OS Triticum aestivum.
XX
PN WO9553068-A2.
XX
PD 21-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07562.
XX
PR 09-APR-1998; 98US-0081148.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Hitz WD, Rafalski JA;
XX
XX WPI; 1999-620432/53.
DR N-PSDB; AA23135.
XX
XX New sucrose transport proteins from plants, useful for controlling
PT carbohydrate transport and distribution in plant cells -
XX
PS Claim 5; Page 56-57; 64pp; English.
XX
XX The invention provides nucleic acid sequences (AA23124-Z23135) encoding
CC sucrose transport proteins (AA41114-Y41125) derived from corn, rice,
CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can
CC be recombinantly expressed by standard recombinant methodology. The
CC invention facilitates studies on carbohydrate metabolism and function in
CC plants, provides genetic tools for the manipulation of these
CC biosynthetic pathways, and provides a means to control carbohydrate
CC transport and distribution in plant cells.
XX
SQ Sequence 563 AA;

Query Match 32.9%; Score 48; DB 20; Length 563;
Best Local Similarity 37.0%; Pred. No. 72;
Matches 10; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 1 VEGELYQGVPRAPGTEARRHYDEGVR 27

```

```

Query Match      32 28:      Score 47: DB
Protein 1:      42.9%:      Prot No 14:
Matches 12:      Conservative 4:      Mismatches 6:      Gaps 1:

QY      3      EGYLGQV-----RAEPGTEARHYDE 24
      ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      42      eglsqglvdrekglsaeppgwakreeee 69

RESULT      16
AAAY53797
XD      AAAY53797 standard; protein; 439 AA.
XC
XC
AC      AAAY53797:

```


PR	22-SEP-1999;	99US-0155139.	PR	09-MAR-1999;	99US-0123548.
PR	23-SEP-1999;	99US-0155486.	PR	23-MAR-1999;	99US-0125788.
PR	24-SEP-1999;	99US-0155659.	PR	25-MAR-1999;	99US-0126264.
PR	28-SEP-1999;	99US-0156458.	PR	29-MAR-1999;	99US-0126785.
PR	29-SEP-1999;	99US-0156596.	PR	01-APR-1999;	99US-0126785.
PR	04-OCT-1999;	99US-0157117.	PR	01-APR-1999;	99US-0126234.
PR	05-OCT-1999;	99US-0157117.	PR	08-APR-1999;	99US-0128714.
PR	05-OCT-1999;	99US-0157125.	PR	16-APR-1999;	99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130890.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0159294.	PR	30-APR-1999;	99US-0132047.
PR	14-OCT-1999;	99US-0159330.	PR	04-MAY-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999;	99US-0132485.
PR	18-OCT-1999;	99US-0159584.	PR	07-MAY-1999;	99US-0132487.
PR	21-OCT-1999;	99US-0160747.	PR	11-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134221.
PR	22-OCT-1999;	99US-0160815.	PR	14-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160980.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160989.	PR	20-MAY-1999;	99US-0135124.
PR	25-OCT-1999;	99US-0161404.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161405.	PR	24-MAY-1999;	99US-0135629.
PR	26-OCT-1999;	99US-0161359.	PR	25-MAY-1999;	99US-0136021.
PR	26-OCT-1999;	99US-0161360.	PR	26-MAY-1999;	99US-0136782.
PR	28-OCT-1999;	99US-0161360.	PR	01-JUN-1999;	99US-0137222.
PR	28-OCT-1999;	99US-0161920.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161992.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161993.	PR	07-JUN-1999;	99US-0137724.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUN-1999;	99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138844.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUN-1999;	99US-0139119.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139453.
PR	29-OCT-1999;	99US-0162142.	PR	17-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139456.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139457.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139458.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139458.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139460.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139462.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139462.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139463.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139750.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139763.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUN-1999;	99US-0139899.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140323.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140323.
PR	29-OCT-1999;	99US-0162142.	PR	24-JUN-1999;	99US-0140695.
PR	29-OCT-1999;	99US-0162142.	PR	24-JUN-1999;	99US-0140823.
PR	29-OCT-1999;	99US-0162142.	PR	25-JUN-1999;	99US-0140823.
PR	29-OCT-1999;	99US-0162142.	PR	25-JUN-1999;	99US-0140991.
PR	29-OCT-1999;	99US-0162142.	PR	30-JUN-1999;	99US-0141287.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0141842.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0141842.
PR	29-OCT-1999;	99US-0162142.	PR	02-JUL-1999;	99US-0142054.
PR	29-OCT-1999;	99US-0162142.	PR	02-JUL-1999;	99US-0142054.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUL-1999;	99US-0142390.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUL-1999;	99US-0142803.
PR	29-OCT-1999;	99US-0162142.	PR	09-JUL-1999;	99US-0142803.
PR	29-OCT-1999;	99US-0162142.	PR	13-JUL-1999;	99US-0142920.
PR	29-OCT-1999;	99US-0162142.	PR	13-JUL-1999;	99US-0142920.
PR	29-OCT-1999;	99US-0162142.	PR	13-JUL-1999;	99US-0142977.
PR	29-OCT-1999;	99US-0162142.	PR	13-JUL-1999;	99US-0143542.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUL-1999;	99US-0143542.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUL-1999;	99US-0143624.

Query Match Similarity 32.28; Score 47; DB 21; Length 913; Indels 0; Gaps 0; Mismatches 6; P ind No 1.7e+02; Matches 9; Conservative 3;

Oy 7 OGVPRAEPTGTEARRHYDE 24
Db 849 rgrprkprkterkrifde 866

RESULT 18
AAG37416
ID AAG37416 standard; Protein; 916 AA.

XX AAG37416;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 46002.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 46002.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.

PR 16-APR-1999; 99US-0129845.
 PR 17-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132007.
 PR 01-MAY-1999; 99US-0132207.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0133119.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135247.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137228.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 14-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139464.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142809.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0143977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144082.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145115.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148111.
 PR 12-AUG-1999; 99US-0148319.
 PR 13-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 16-AUG-1999; 99US-0148684.
 PR 17-AUG-1999; 99US-0149368.
 PR 18-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 23-AUG-1999; 99US-0150566.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155569.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157752.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.

```
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 32.2%; Score 47; DB 21; Length 994;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 7 OGVPRAEPTGEARRHYDE 24
Db 930 rgrprkprcterkrfide 947

RESULT 20
AAG37415
ID AAG37415 standard; Protein; 997 AA.
XX AC AAG37415;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 45001.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132046.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135829.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
```

```
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145386.
PR 26-JUL-1999; 99US-0145274.
PR 26-JUL-1999; 99US-0145274.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146386.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147360.
PR 06-AUG-1999; 99US-0147360.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0152750.
PR 13-SEP-1999; 99US-0152750.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156259.
PR 28-SEP-1999; 99US-0156259.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158272.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
Query Match 21 22.28; Score 47; DB 21; Length 997;
Best Local Similarity 30.08; Predicted No. 1.9e+02;
Matches 3; Mismatches 6; Indels 0; Gaps 0;
QY 7 QGVPRAPGTEARRHYDE 24
DB 933 rgrprkrpeterkrifde 950
RESULT 21
AAG14136
ID AAG14136 standard; Protein: 1098 AA.
XX
AC AAG14136;
XX
XX 17-OCT-2000 (first-entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 13885.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126785.
PR 23-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130440.
PR 21-APR-1999; 99US-0130440.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132482.
PR 06-MAY-1999; 99US-0132482.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
```


PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 14-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136382.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 11-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 23-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.


```
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 32.2%; Score 47; DB 21; Length 1101;
Best Local Similarity 50.0%; Pred. No. 2.1e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 7 QGVPRAEPTTEARRHYDE 24
Db 1037 rgrprkpeterkrlfde 1054
:| || | || :| ||
:| || | || :| ||

RESULT 23
AAW42074
ID AAW42074 standard; Protein; 140 AA.
XX
AC AAW42074;
DT 07-JUL-1998 (first entry)
XX
DE The amino acid sequence of the C195_1 protein.
XX
KW C195_1 protein; human PBMC cDNA library; nutritional activity;
KW peripheral blood mononuclear cell; nutritional activity;
KW cytokine activity; cell proliferation/differentiation activity;
KW homology; EST; expressed sequence tag.
XX
OS Homo sapiens.
XX
PN WO9748801-A2.
PD 24-DEC-1997.
XX
PF 16-JUN-1997; 97WO-US10501.
PR 17-JUN-1996; 96US-0664596.
XX
PA (GEM ) GENETICS INST INC.
XX
PI Bowman M, Evans C, Jacobs K, Lavallie ER, McCoy JM;
PI Merberg D, Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1998-063142/06.
DR N-PSDB; AAV09268.
XX
PT Poly:nucleotide(s) and proteins obtained from human PBMC, dendritic
PT cell, adult brain, fetal brain and adult testes cDNA libraries -
PT used in research, detection and therapy of, e.g. cytokine and cell
PT proliferation or differentiation
XX
PS Claim 1; Pages 47-48; 78pp; English.
XX
CC This is the amino acid sequence of the C195_1 protein which was
CC isolated from a human PBMC cDNA library. It can be used in research,
CC detection and therapy, as they may have nutritional activity, cytokine
CC and cell proliferation/differentiation activity. A search of the
CC sequence encoding this protein against the Genbank database
CC demonstrated that this sequence has at least some homology with two ESTs
CC identified as "yg11g06.r1 Homo sapiens cDNA clone 31937 5'" and
CC "ym13f12.r1 Homo sapiens cDNA clone 48025 5'".
XX
SQ Sequence 140 AA;

Query Match 31.8%; Score 46.5; DB 19; Length 140;
Best Local Similarity 47.4%; Pred. No. 26;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYOGVPRAEPTTEARRHYD 23
Db 83 vwsqvpampgpk-rhhd 100
:: ||| ||| :| ||
```

RESULT 24
AAW41600
ID AAW41600 standard; Protein: 140 AA.
XX AC AAW41600;
XX AC AAW41601;
DT 22-JUN-1998 (first entry)
XX DE Secreted protein C195-1.
XX DE Secreted protein C195-1.
KW Secreted protein; C195-1; peripheral blood mononuclear cell; PBMC;
KW protein factor; human.
XX OS Homo sapiens.
XX FH Key
XX FH Protein
XX FT Location/Qualifiers
XX FT 52..140
XX FT /note= "Claim 10"
XX WO9748800-A1.
XX 24-DEC-1997.
XX 16-JUN-1997; 97WO-US10500.
XX 25-OCT-1996; 96US-0738367.
XX 17-JUN-1996; 96US-0664596.
XX 27-SEP-1996; 96US-0721926.
XX (GEMY) GENETICS INST INC.
XX Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA;
XX Spaulding V, Treacy R;
XX N-PSDB; AAV04273.
XX Nucleic acid encoding secreted protein from human peripheral blood
XX mononuclear cells - useful, e.g. as immunomodulators, antitumor
XX agents, promoters of tissue growth, haemostatic and thrombolytic
XX agents etc.
XX Claim 8; Page 35-36; 52pp; English.
XX This polypeptide, designated C195_1, comprises a novel secreted
XX protein of human peripheral blood mononuclear cells (PBMC). Its
XX amino acid sequence was deduced from the cDNA insert (see AAV04273)
XX of isolate C195_1 deposited under accession number ATCC 98079.
XX Claimed host cells can be used in the recombinant production of
XX C195_1 protein. Claimed compositions including the protein or its
XX fragments can be used in a claimed method for preventing, treating
XX or ameliorating a medical condition. The secreted protein may also
XX be used to screen compounds for biological activity; to raise
XX antibodies, and as a nutritional source. It may also have
XX biological activities, e.g. cytokine, cell proliferation or
XX differentiation activity, immunosuppressant, immunostimulant,
XX regulation of haematopoiesis, modulation of fertility, chemotactic,
XX chemokinetic, haemostatic, thrombolytic, antimicrobials, analgesic,
XX antipsoriatic, etc. No evidence is given to support any of these
XX activities or applications.
XX Sequence 140 AA;

Query Match 31.8%; Score 46.5; DB 19; Length 140;
Best Local Similarity 47.4%; Pred. No. 26;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;
OY 5 LYQGVPRAEPTGTEARRHYD 23
::: ||| ||| | | |

Db 83 vwsyvpaempgk-rhhfd 100
RESULT 25
AAW41601
ID AAW41601 standard; Protein: 240 AA.
XX AC AAW41601;
DT 22-JUN-1998 (first entry)
XX DE Secreted protein C195-4.
XX DE Secreted protein; C195-4; peripheral blood mononuclear cell; PBMC;
KW protein factor; human.
XX OS Homo sapiens.
XX FH Key
XX FH Protein
XX FT Location/Qualifiers
XX FT 1..140
XX FT /note= "Claim 34"
XX WO9748800-A1.
XX 24-DEC-1997.
XX 16-JUN-1997; 97WO-US10500.
XX 25-OCT-1996; 96US-0738367.
XX 17-JUN-1996; 96US-0664596.
XX 27-SEP-1996; 96US-0721926.
XX (GEMY) GENETICS INST INC.
XX Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA;
XX Spaulding V, Treacy R;
XX N-PSDB; AAV04274.
XX Nucleic acid encoding secreted protein from human peripheral blood
XX mononuclear cells - useful, e.g. as immunomodulators, antitumor
XX agents, promoters of tissue growth, haemostatic and thrombolytic
XX agents etc.
XX Claim 8; Page 38-39; 52pp; English.
XX This polypeptide, designated C195_4, comprises a novel secreted
XX protein of human peripheral blood mononuclear cells (PBMC). Its
XX amino acid sequence was deduced from the cDNA insert (see AAV04274)
XX of isolate C195_4 deposited under accession number ATCC 98192.
XX Claimed host cells can be used in the recombinant production of
XX C195_4 protein. Claimed compositions including the protein or its
XX fragments can be used in a claimed method for preventing, treating
XX or ameliorating a medical condition. The secreted protein may also
XX be used to screen compounds for biological activity; to raise
XX antibodies, and as a nutritional source. It may also have
XX biological activities, e.g. cytokine, cell proliferation or
XX differentiation activity, immunosuppressant, immunostimulant,
XX regulation of haematopoiesis, modulation of fertility, chemotactic,
XX chemokinetic, haemostatic, thrombolytic, antimicrobials, analgesic,
XX antipsoriatic, etc. No evidence is given to support any of these
XX activities or applications.
XX Sequence 240 AA;

Query Match 31.8%; Score 46.5; DB 19; Length 240;
Best Local Similarity 47.4%; Pred. No. 47;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;
OY 5 LYQGVPRAEPTGTEARRHYD 23

XX SR-p70: monkey; transcription factor; p53; tumour suppressor gene;
 KW homology; differential splicing; diagnosis; cancer; neuroblastoma;
 KW gene therapy; apoptosis.
 OS Cercopithecus aethiops.
 XX WO9728186-A1.
 YN 07-AUG-1997.
 PD
 XX 03-FEB-1997; 97WO-FR00214.
 PF
 XX 02-FEB-1996; 96FR-0001309.
 XX
 XX (SNFI) SANOFI SA.
 XX Caput D, Ferrara P, Kaghad AM;
 XX WPI; 1997-402550/37.
 DR N-PSDB; AAV01497.
 DR New polypeptide(s) encoded by the SR-p70 tumour suppressor gene -
 XX and related nucleic acid, useful for diagnosis and treatment of
 XX tumours
 XX
 XX Claim 1; Fig 5; 136pp; French.
 PS This is the amino acid sequence of the protein SR-p70b from monkey
 XX of p53-regulated genes and are expressed by tumour suppressor genes
 CC related to the p53 gene family. The gene sequence was isolated from a
 CC cDNA library by sequencing the inserts and comparing to sequence
 CC databases. The protein sequence contains regions of homology to the p53
 CC protein. The SR-p70b gene sequence was isolated simultaneously with the
 CC SR-p70a sequence (AAV01497) from the same monkey. The sequences can be
 CC used in the diagnosis and monitoring of cancer, especially neuroblastoma.
 CC The nucleic acid sequences and corresponding antisense sequences, are
 CC also useful in gene therapy, e.g. to regulate apoptosis.
 XX
 XX Sequence 499 AA;
 SQ
 Query Match 31.5%; Score 46; DB 18; Length 499;
 Best Local Similarity 56.2%; Pred. No. 1.2e+02;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;
 OY 9 VPRAEPTTEARRHYDE 24
 DE | | | | |
 DB 337 vpalgpyvkkrrhgde 352
 RESULT 29
 AAR72328
 ID AAR72328 standard; Protein: 529 AA.
 XX
 XX AAR72328;
 XX
 XX 14-OCT-1995 (first entry)
 DT
 XX Laccase Rslac3.
 DE Rslac3; laccase; lignin; lignosulfonate; polymerization;
 XX depolymerization.
 KW
 XX Rhizoctonia solani.
 OS
 XX WO9507988-A.
 PN
 XX 23-MAR-1995.
 XX
 XX 13-SEP-1994; 94WO-US10264.
 PF

XX 17-SEP-1993; 93US-0122230.
 PR 17-SEP-1993; 93US-0122827.
 PR 03-DEC-1993; 93US-0162827.
 PR 22-DEC-1993; 93US-0172331.
 XX
 XX (NOVO) NOVO-NORDISK BIOTECH INC.
 EA
 XX (NOVO) NOVO-NORDISK AS.
 XX Christensen BE, Schneider P, Wahleithner JA;
 PI WPI; 1995-131356/17.
 XX N-PSDB; AA086526.
 DR
 XX New neutral Rhizoctonia laccase(s) and corresp. nucleic acids -
 XX are used industrially for polymerising lignin, lignosulphate(s)
 PT and phenolic cpds. and for oxidising dyes.
 PT
 XX Claim 6; Page 44-46; 78pp; English.
 PS
 CC A. R. solani RS22 (IMI CC 358730) cDNA library was screened on
 CC ABTS. The DNA insert of a plasmid from an isolated clone was
 CC sequenced to reveal a novel gene, Rslac3 (given in AA086526),
 CC encoding a laccase (AAR72328) optimally active at pH 6.0-8.5.
 CC Recombinant laccase was expressed in E. coli.
 XX
 XX Sequence 529 AA;
 SQ
 Query Match 31.5%; Score 46; DB 16; Length 529;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 OY 6 YOGVPRAEPTTE 17
 DE | | | | |
 DB 314 Yegapnaeptte 325
 RESULT 30
 AAW16301
 ID AAW16301 standard; Protein: 529 AA.
 XX
 XX AAW16301;
 XX
 XX 07-AUG-1997 (first entry)
 DT
 XX Rhizoctonia solani laccase isozyme 4.
 DE
 XX Blue copper oxidase; laccase; enzyme engineering;
 KW protein engineering; lignin depolymerisation; dye oxidation.
 XX
 XX Rhizoctonia solani.
 OS
 XX
 XX Key Location/Qualifiers
 FH 427
 FT Binding-site /note= "Type I copper site ligand"
 FT Binding-site 470
 FT Binding-site /note= "possible Type I copper site ligand"
 FT Binding-site 480
 FT Binding-site /note= "Type I copper site ligand"
 FT Binding-site 480
 FT Peptide /note= "Type I copper site ligand"
 FT Peptide 217..226
 FT /note= "preferred site for mutation (Claim 30),
 FT corresponds to W163 region of ascorbate
 FT oxidase"
 FT Peptide 303..312
 FT /note= "preferred site for mutation (Claim 29),
 FT corresponds to R285 region of ascorbate
 FT oxidase"
 FT Peptide 356..371
 FT /note= "preferred site for mutation (Claim 28),
 FT corresponds to Q353 and W362 region of

FT	Peptide	416..429	ascorbate oxidase"
FT	/note=	"preferred site for mutation (Claim 26),	
FT		contains Cu-ligating His"	
FT	Peptide	465..472	
FT	/note=	"preferred site for mutation (Claim 31)"	
FT	Peptide	466..470	
FT	/note=	"preferred site for mutation (Claim 18)"	
FT	Peptide	474..494	
FT	/note=	"preferred site for mutation (Claim 27),	
FT		contains Cu-ligating His and Cys"	
FT	Misc-difference	379	
FT	/note=	"translated residue from gene sequence is	
FT		Ile"	
FT	Misc-difference	415	
FT	/note=	"translated residue from gene sequence is	
FT		Ile"	
XX			
PN	WO9709431-A1.		
XX			
XX	13-MAR-1997.		
XX			
XX	03-SEP-1996;	96WO-US14087.	
PF			
XX			
XX	01-SEP-1995;	95US-0003142.	
PR			
XX			
PA	(NOVO) NOVO NORDISK BIOTECH INC.		
XX			
PI	Berka RM, Wahleithner JA, Xu F, Berka R;		
XX			
DR	WPI; 1997-129906/17.		
DR	N-PSDB; AAT63317.		
XX			
XX			
PT	New mutant blue copper oxidase enzymes - having different specific		
PT	activities to wild-type enzymes, used for e.g. lignin		
PT	(de)polymerisation or oxidation of dyes		
XX			
PS	Disclosure; Fig 6A-D; 48pp; English.		
XX			
CC	Rhizoctonia solani laccase isozyme 4 (AAW16301) can be engineered		
CC	alter e.g. its specific activity or pH-activity profile or to		
CC	improve expression yields. The mutant laccase is prepd. by site-		
CC	directed mutagenesis of the rs14 gene (AAT63317) and expression in		
CC	host cells. The mutation may be a deletion, insertion or pref. a		
CC	substn. of one or more amino acids at a location no greater than		
CC	12 Angstroms, pref. no more than 2.5 Angstroms, from the Type I		
CC	copper site. The mutant enzyme can be used e.g. for the		
CC	polymerisation or depolymerisation of lignins, oxidation of dyes,		
CC	or polymerisation of phenolic or aniline cpds. present in liquids,		
CC	e.g. for the treatment of fruit juices.		
XX			
SQ	Sequence	529 AA;	
	Query Match	31.5%; Score 46; DB 18;	
	Best Local Similarity	66.7%; Pred. No. 1.3e+02;	
	Matches	8; Conservative 1; Mismatches 3; Indels 0; Gaps	
Qy	6 YQGVPRAPQTE 17		
	I : I I I I I I		
Db	314 yegapnaeptte 325		
RESULT	31		
ID	AAW76310		
XX	AAW76310 standard; protein; 529 AA.		
AC			
XX	AAW76310;		
XX			
DT	08-JAN-1999 (first entry)		
XX			
DE	Rhizoctonia solani (I) laccase protein.		
XX			

KW	Laccase; variant; oxidation; dye transfer inhibition; bleaching;
KW	denim; lignin modification; paper strengthening; phenol polymerisation;
KW	hair dye; waste water treatment.
XX	
OS	Rhizoctonia solani.
XX	
PN	WO9838287-A1.
XX	
PD	03-SEP-1998.
XX	
XX	23-FEB-1998; 98WO-DK00070.
PF	
XX	
PR	28-FEB-1997; 97DK-0000222.
XX	
PA	(NOVO) NOVO-NORDISK AS.
XX	
PI	Svendsen A, Xu F;
XX	
DR	WPI; 1998-495393/42.
XX	
PT	New variants of Coprinus and related laccases with increased
PT	oxidation potential - or altered pH optimum, or mediator or
PT	oxygen-hydroxide ion pathways, useful for oxidation, for inhibiting
PT	dye transfer and in bleaching textiles, especially as detergent
PT	additive
XX	
PS	Disclosure; Pages 124-125; 147pp; English.
XX	
CC	The present sequence represents a laccase protein. The specification
CC	describes active laccase variants (see AAW76282, AAW76296-99 and
CC	AAW76316-17) having increased oxidation potential, altered pH optimum,
CC	altered mediator and/or altered oxygen/hydroxide ion pathway. The
CC	laccase variants are used specifically to oxidise substrates, to
CC	inhibit dye transfer, and for bleaching textiles, specifically denim.
CC	They can also be used for lignin modification, strengthening paper,
CC	polymerisation of phenols, dyeing of hair and textiles and waste
CC	water treatment.
XX	
SQ	Sequence 529 AA;
Query Match 31.5%; Score 46; DB 19; Length 529;	
Best Local Similarity 66.7%; Pred. No. 1.3e+02;	
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	
QY	6 YQGVPRAPCTE 17
:	
DB	314 yegapnaepte 325
RESULT 32	
ID	AAW60878 standard; Protein; 529 AA.
XX	
AC	AAW60878;
XX	
DT	09-NOV-1998 (first entry)
XX	
DE	Rhizoctonia solani (I) laccase.
XX	
KW	Coprinus laccase-like enzyme; enzyme engineering; enzyme stability;
KW	detergent; bleaching.
XX	
OS	Rhizoctonia solani.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 126
FT	/note- *variant may have Ala, Val, Leu, Ile, Pro,
FT	Phe, Met, Gly, Ser, Thr, Cys, Tyr, Asn,
FT	Gln, Asp, Glu, Lys, Arg or His (preferably
FT	Phe or His) at this position"
FT	Misc-difference 135
FT	/note- *variant may have Ala, Val, Leu, Ile, Pro,

DT 27-NOV-2000 (first entry)
DE Eucalyptus grandis cell signalling involved protein SEQ ID NO:913.
XX
XX Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
KW plant cell signalling; modulation; transgenic plant; pathogen; growth;
KW environmental change; development; cell proliferation; differentiation;
XX elongation; survival; disease resistance; nutrient metabolism.
OS Eucalyptus grandis.
XX
XX WO200042171-A1.
PN 20-JUL-2000.
XX
XX 11-JAN-2000; 2000WO-US00724.
PF
XX 12-JAN-1999; 99US-0228986.
PR 01-NOV-1999; 99US-0162866.
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA Strabala TJ, Nieuwenhuizen NJ;
PI WPI; 2000-476052/41.
DR
XX
XX Isolated polynucleotide encoding a polypeptide involved in cell
PT signaling used for generating transgenic plants with modified responses
PT to external signals -
XX
XX Claim 3; Page 475-476; 527pp; English.
XX
XX AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
CC pine (Pinus radiata) also known as Monterey pine). The protein sequences
CC are involved in cell signalling. The polynucleotide and protein
CC sequences can be used to modify the response of plant cells to external
CC signals e.g. environmental changes or pathogens during the growth and
CC development of a plant. They can be used to modify cell proliferation,
CC differentiation, elongation and survival, resistance to disease and
CC nutrient metabolism. Examples of modifications which can be produced are
CC altered fruit ripening and senescence of leaves and flowers e.g. to
CC delay senescence and prolong the life of cut flowers or enhance
CC senescence of reproductive organs to engineer sterile plants. Other
CC modifications can be used to delay senescence in selected cell types or
CC organs providing fruit and vegetables which have a longer shelf life
CC between harvest and consumption, or to decrease branching frequency in
CC forest tree species giving long stretches of valuable knot-free clear
CC wood which can be used in solid timber furniture and veneers.
XX
XX Sequence 541 AA;
SQ
Query Match 31.5%; Score 46; DB 21; Length 541;
Best Local Similarity 40.9%; Pred. No. 1.3e+02;
Matches 9; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
Oy 5 LYQGVPRAEPTGTEARRHYDEGV 26
|||
Db 237 lyggipetisdmealehldvgi 258
|||
RESULT 35
AAW86348
ID AAW86348 standard; Protein; 566 AA.
XX
AC AAW86348;
XX
XX 15-MAR-1999 (first entry)
DT Human multiple endocrine receptor neoplasia type 1 protein.
DE Human; multiple endocrine neoplasia type 1 protein; MEND-1;
XX
KW

KW tumorigenesis; cancer; multiple endocrine neoplasia; gene therapy;
KW detection; diagnosis; drug screening.
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FH Misc-difference 137 /label= unknown
FT /note= "encoded by YGG"
FT Misc-difference 161 /label= unknown
FT /note= "encoded by NNT"
FT Misc-difference 274 /label= unknown
FT /note= "encoded by RGG"
FT Misc-difference 312 /label= unknown
FT /note= "encoded by SCC"
FT Misc-difference 417 /label= unknown
FT /note= "encoded by TNG"
FT Misc-difference 421 /label= unknown
FT /note= "encoded by NGC"
FT Misc-difference 424 /label= unknown
FT /note= "encoded by GAN"
FT Misc-difference 428 /label= unknown
FT /note= "encoded by CNG"
FT Misc-difference 435 /label= unknown
FT /note= "encoded by ANC"
FT Misc-difference 443 /label= unknown
FT /note= "encoded by NCC"
XX WO9854324-A1.
XX
XX 03-DEC-1998.
XX
XX 29-MAY-1998; 98WO-US10957.
XX
XX 29-MAY-1997; 97US-0865337.
XX (INCY-) INCYTE PHARM INC.
PI Au-Young J, Covitz PA, Murry LE, Tang YT;
XX WPI; 1999-059839/05.
DR N-PSDB; AAV80659.
XX
XX New isolated multiple endocrine neoplasia type 1 protein - used to
PT develop products for the diagnosis, treatment and prevention of
PT cancers and multiple endocrine neoplasia.
XX
XX Claim 1; Fig 1; 67pp; English.
XX
XX The present sequence is human multiple endocrine neoplasia type 1
CC protein (MEND-1). MEND-1 plays a role in multiple endocrine neoplasia
CC when one or both normal genetic copies of MEND-1 are mutated and no
CC longer able to suppress tumorigenesis. MEND-1 can be used for treating
CC or preventing cancers and multiple endocrine neoplasia. MEND-1
CC polynucleotides can also be used for gene therapy. Products from the
CC present invention can also be used for detection, diagnosis and drug
CC screening.
XX
XX Sequence 566 AA;
SQ
Query Match 31.5%; Score 46; DB 20; Length 566;
Best Local Similarity 52.9%; Pred. NO. 1.4e+02;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 2 GEGLYOCVPRAEPTGA 18
 Db 507 gggavsgpprkppgtva 523

RESULT 37
 ID AAB40554 standard; Protein; 619 AA.
 AC AAB40554;
 DT 08-FEB-2001 (first entry)
 DE Human ORF318 polypeptide sequence SEQ ID NO:636.
 KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; neurotrophic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunosuppressant; thrombotic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineoplastic; antithyroid;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease; SCID;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; rhythematosus; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 OS Homo sapiens.
 PN WO200058473-A2.
 XX 05-OCT-2000.
 XX 31-MAR-2000; 2000WO-US08621.
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 XX 30-MAR-2000; 2000US-0340763.
 XX (CURA-) CURAGEN CORP.
 PA Shinkets RA, Leach M;
 PI WPI: 2000-502362/57.
 CC N-PSDB; AAC4763.
 DR Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 CC neurodegenerative disorders and cardiovascular disease -
 XX Claim 11; Page 760-761; 5507pp; English.
 CC AAC47446 to AAC77606 encode the proteins given in AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; neurotrophic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antifungal; antineoplastic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC applications. The proteins can be used to treat the following disorders:
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,

OY 2 GEGLYOCVPRAEPTGA 18
 Db 463 gggavsgpprkppgtva 479

RESULT 36
 ID AAW29749 standard; Protein; 610 AA.
 AC AAW29749;
 DT 11-JAN-1999 (first entry)
 DE Tumor suppressor protein menin.
 KW Familial multiple endocrine neoplasia type 1; FMEN1; MEN1;
 KW menin; tumor suppressor gene; cancer; marker; diagnosis;
 KW gene therapy; human.
 XX Homo sapiens.
 PN WO9839439-A1.
 XX 11-SEP-1998.
 XX 04-MAR-1998; 98WO-US04258.
 PR 05-MAR-1997; 97US-0040269.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Agarwal SK, Burns AL, Chandrasekharappa SC, Collins FS;
 PI Debelenko LV, Emmert-Buck MR, Guru SC, Liotta LA;
 PI Lubensky IA, Manickam P, Marx SJ, Spiegel AM;
 DR WPI: 1998-506360/43.
 DR N-PSDB; RAV57415-16.
 XX Protein and gene associated with multiple endocrine neoplasia type 1
 XX - useful in gene therapy and to diagnose sufferers of, and those
 XX susceptible to, this condition by detecting protein absence or gene
 XX mutation(s)
 XX Claim 11; Page 56-57; 75pp; English.
 XX This is the amino acid sequence of menin, a 67.5 kDa polypeptide
 XX encoded by a novel human tumor suppressor gene, MEN1 (see AAV27416),
 XX which exhibits no apparent amino acid sequence similarities to previously
 XX known proteins. The lack of a functional menin polypeptide, either
 XX by absence of the protein, its alteration and/or associated
 XX mutations in the MEN1 gene, have been identified in individuals
 XX with familial multiple endocrine neoplasia type 1 (FMEN1) and
 XX identified from multiple endocrine neoplasia type 1. The mechanism
 XX of endocrine tumorigenesis, facilitates accurate early diagnosis of
 XX MEN1 associated cancers, and provides preclinical identification of
 XX individuals with the FMEN1 syndrome, i.e. cancer-free individuals
 XX that are at high risk of acquiring MEN1 associated tumours. It
 XX also provides a basis for gene therapy. Menin can be expressed in
 XX suptic cells, prokaryotic cells and meskaryotic cells. Menin
 XX can be used to identify menin
 XX humans cells or tissues, especially from patients suspected of
 XX being at risk from multiple endocrine neoplasia type 1.
 XX Sequence 610 AA;
 XX Query Match 31.5%; Score 46; DB 19; Length 610;
 XX Best Local Similarity 52.9%; Pred. No. 1.5e+02;
 XX Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Best local similarity	50.2%	Seq. NO. 1.0e+02,
Matches	9;	Mismatches
Conservative	1;	Indels
Gaps	0;	Gaps

OY 6 YGVPRAPGTEARRH 21
 Db 78 ypgipgrpgsawrh 93

RESULT 40

AA52702 standard; Protein; 1277 AA.
 AC AAR52702;
 DT 11-JAN-1995 (first entry)
 XX Plasmid pASK60-Strep reading frame "c" translation.
 XX Streptavidin binding peptide; fusion protein: pASK60-Strep;
 KW affinity chromatography; purification; peptide tag; detection.
 XX Synthetic.
 OS
 XX Key Location/Qualifiers
 XX Region 1-1277
 FT /note- "translated from reading frame 'c' of
 FT pASK60-Strep; all X's correspond to nonsense
 FT Codons"
 FT Peptide 482..491
 XX /label- streptavidin-binding peptide
 PN GB2272698-A.
 XX
 XX 25-MAY-1994.
 XX
 XX 01-NOV-1993; 93GB-0022501.
 XX
 XX 03-NOV-1992; 92DE-4237113.
 XX
 XX (KUEH-) KUEHN KONSTRUKTION GMBH & CO KG KLAUS.
 XX (BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.
 XX
 XX Schmidt T. Skerra A;
 PI WPI: 1994-153484/19.
 DR N-PSDB; AAO62676.
 DR
 XX New fusion peptide(s) - have easily controlled binding properties
 PT and are capable of binding to streptavidin
 XX
 XX Example 5; Fig 8; 53pp; English.
 XX
 CC Plasmid pASK60-Strep was produced starting from pASK40
 CC (Biotechnology 9, 273-278, 1991) using site-directed mutagenesis
 CC and PCR. The plasmid contains an improved set of unique restriction
 CC sites, including two sites located directly at the 3' end of the
 CC region coding for the ompA signal peptide. The polylinker is
 CC located between the ompA signal peptide and the streptavidin binding
 CC peptide. The plasmid pASK60-Strep is useful for the expression of
 CC polypeptides fused to a streptavidin binding peptide. The
 CC production of such fusion proteins can be specifically detected
 CC using a streptavidin-alkaline phosphatase conjugate.
 XX
 XX Sequence 1277 AA;

Query Match 31.5%; Score 46; DB 15; Length 1277;
 Best Local Similarity 50.0%; Pred. No. 3.4e+02;
 Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 6 YGVPRAPGTEARRH 21
 Db 471 ypgipgrpgsawrh 486

RESULT 41

AA52702 standard; Protein; 170 AA.
 AC AAB36228;
 DT 19-FEB-2001 (first entry)
 XX Rat Fas receptor.
 XX Rat Fas receptor.
 KW Rat; Fas receptor; gene therapy; vascular cell; stenosis; FGFR-1;
 KW accelerated graft arteriosclerosis; vascular occlusive disease.
 XX Rattus sp.
 OS
 XX WO200063369-A2.
 XX
 XX 26-OCT-2000.
 XX
 XX 18-APR-2000; 2000WO-EP03532.
 XX
 XX 19-APR-1999; 99US-0294176.
 XX 10-MAY-1999; 99US-0307690.
 XX
 XX (NOVS) NOVARTIS AG.
 XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX (UYJO) UNIV JOHNS HOPKINS.
 XX Ballermann BJ, Goldschmidt P, Movva RN, Sanfilippo A;
 XX WPI: 2000-672735/65.
 XX N-PSDB; AAC66557.
 XX
 XX Genetic modification of target cells for inhibiting excessive
 XX proliferation, comprises transferring, to cells or progenitors, the DNA
 XX sequence encoding Fas antigen or receptor, or the fibroblast growth
 XX factor receptor.
 XX
 XX Disclosure; Page 33-34; 41pp; English.
 XX
 XX The present invention describes a method of gene therapy which involves
 XX the use of Rask or FGFR1 coding sequences to treat vascular diseases.
 XX This method includes the use of Rask or FGFR1 coding sequences to treat
 XX diseases include vascular occlusive diseases associated with cell
 XX proliferation, such as accelerated graft arteriosclerosis and other forms
 XX of stenosis.
 XX
 XX Sequence 170 AA;

Query Match 30.8%; Score 45; DB 21; Length 170;
 Best Local Similarity 26.4%; Pred. No. 53;
 Matches 14; Conservative 2; Mismatches 9; Indels 28; Gaps 2;
 OY 3 EGIYQ-----GVPRAPGTEA-----RHVDEGVR 27
 Db 46 eglyqgpcqpcqgervkvdttsgaptchpcgeeytdrkhyadkcr 98

RESULT 42

AA52702 standard; Protein; 411 AA.
 AC AAY87981;
 DT 21-SEP-2000 (first entry)
 XX Pseudomonas sp. HR199 vdh-delta protein.
 DE Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
 KW vanillic acid; flavouring; vanillin dehydrogenase; vdh;
 XX vdh-delta.

OS Pseudomonas sp.
 XX DE19850242-A1.
 PN
 XX
 XX
 PD 04-MAY-2000.
 XX
 PF 31-OCT-1998; 98DE-1050242.
 XX
 PR 31-OCT-1998; 98DE-1050242.
 XX
 PA (HAAR) HAARMANN & REIMER GMBH.
 XX
 PI Rabenhorst J, Steinbuechel A, Priefert H, Overhage J;
 XX WPI: 2000-340642/30.
 XX DR N-PSDB; AAA39731.
 XX
 PT Transformed organisms, used for production of vanillin and other
 PT methoxyphenols, have altered catabolism of eugenol or ferulic acid
 XX
 PS Disclosure; Fig 2r; 80pp; German.
 XX
 CC This invention describes novel transformed and/or mutated uni- or
 CC multi-cellular organisms (A) in which enzymes (I) involved in catabolism
 CC of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way
 CC that the organism accumulates the intermediates coniferyl alcohol (CA1),
 CC coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
 CC INDEPENDENT CLAIMS are also included for the following: (a) gene
 CC structures in which sequences that encode one or more of the enzymes
 CC CA1-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA
 CC hydratase-aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase
 CC have been altered and/or inactivated; (b) vector containing at least one
 CC gene structure of (a); (c) biotechnical production of organic compounds
 CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
 CC production method of (A). (A) are used to prepare organic compounds,
 CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
 CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
 CC The inexpensive starting material eugenol can be converted to V in a
 CC single step. This sequence represents the Pseudomonas sp. HR199 vanillin
 CC dehydrogenase, vdh-delta which is described in the method of the
 CC invention.
 XX Sequence 411 AA;
 SQ
 Query Match 30.8%; Score 45; DB 21; Length 411;
 Best Local Similarity 47.1%; Pred. No. 1.4e+02;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 4 GLYQGVPRAPGTEARR 20
 ||| |||:: | :|
 Db 81 glysgvpkspalgvqr 97
 RESULT 43
 AAY87980
 ID AAY87980 standard; Protein; 586 AA.
 XX
 AC AAY87980;
 XX
 XX 21-SEP-2000 (first entry)
 DT
 DE Pseudomonas sp. HR199 vdh-omega-Gm protein.
 XX
 XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
 KW vanillic acid; flavouring; vanillin dehydrogenase; vdh;
 KW vdh-omega-Gm.
 XX
 OS Pseudomonas sp.
 XX
 PN DE19850242-A1.
 XX
 XX 04-MAY-2000.

XX 31-OCT-1998; 98DE-1050242.
 XX
 PR 31-OCT-1998; 98DE-1050242.
 XX
 PA (HAAR) HAARMANN & REIMER GMBH.
 XX
 PI Rabenhorst J, Steinbuechel A, Priefert H, Overhage J;
 XX WPI: 2000-340642/30.
 XX DR N-PSDB; AAA39730.
 XX
 PT Transformed organisms, used for production of vanillin and other
 PT methoxyphenols, have altered catabolism of eugenol or ferulic acid
 XX
 PS Disclosure; Fig 2q; 80pp; German.
 XX
 CC This invention describes novel transformed and/or mutated uni- or
 CC multi-cellular organisms (A) in which enzymes (I) involved in catabolism
 CC of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way
 CC that the organism accumulates the intermediates coniferyl alcohol (CA1),
 CC coniferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
 CC INDEPENDENT CLAIMS are also included for the following: (a) gene
 CC structures in which sequences that encode one or more of the enzymes
 CC CA1-dehydrogenase, CA2-dehydrogenase, FA-CoA-synthetase, enoyl-CoA
 CC hydratase-aldolase, beta-ketothiolase, V-dehydrogenase and VA-demethylase
 CC have been altered and/or inactivated; (b) vector containing at least one
 CC gene structure of (a); (c) biotechnical production of organic compounds
 CC (particularly alcohols, aldehydes or organic acids) using (A); and (d)
 CC production method of (A). (A) are used to prepare organic compounds,
 CC preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
 CC FA, V and/or VA, most particularly V, a well-known natural flavouring.
 CC The inexpensive starting material eugenol can be converted to V in a
 CC single step. This sequence represents the Pseudomonas sp. HR199 vanillin
 CC dehydrogenase, vdh-omega-Gm which is described in the method of the
 CC invention.
 XX Sequence 586 AA;
 SQ
 Query Match 30.8%; Score 45; DB 21; Length 586;
 Best Local Similarity 47.1%; Pred. No. 2e+02;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 4 GLYQGVPRAPGTEARR 20
 ||| |||:: | :|
 Db 81 glysgvpkspalgvqr 97
 RESULT 44
 AAY87979
 ID AAY87979 standard; Protein; 673 AA.
 XX
 AC AAY87979;
 XX
 XX 21-SEP-2000 (first entry)
 DT
 DE Pseudomonas sp. HR199 vdh-omega-Km protein.
 XX
 XX Eugenol; ferulic acid; coniferyl alcohol; coniferyl aldehyde; vanillin;
 KW vanillic acid; flavouring; vanillin dehydrogenase; vdh;
 KW vdh-omega-Km.
 XX
 OS Pseudomonas sp.
 XX
 PN DE19850242-A1.
 XX
 PD 04-MAY-2000.
 XX
 XX 31-OCT-1998; 98DE-1050242.
 PF
 XX 31-OCT-1998; 98DE-1050242.
 PR
 XX

PA (HAAR) HAARMANN & REIMER GMBH.
 PI Rabenhorst J, Steinbuechel A, Priefert H, Overhage J;
 XX
 XX WPI: 2000-340642/30.
 DR N-PSDB; AAB87979.
 XX
 XX Transformed organisms, used for production of vanillin and other
 XX methoxyphenols, have altered catabolism of eugenol or ferulic acid
 XX
 XX Disclosure; Fig 2p; 80pp; German.
 XX
 XX This invention describes novel transformed and/or mutated uni- or
 XX cell strains, which are used for the production of vanillin and/or
 XX of eugenol (E) and/or ferulic acid (FA) are inactivated in such a way
 XX that the organism accumulates the intermediates conferyl alcohol (CA1),
 XX conferyl aldehyde (CA2), FA, vanillin (V) and/or vanillic acid (VA).
 XX INDEPENDENT CLAIMS are also included for the following: (a) gene
 XX structures in which sequences that encode one or more of the enzymes
 XX CA1-dehydrogenase, CA2-dehydrogenase, FA-coh-synthetase, and/or CA2-
 XX dehydrogenase have been altered and/or inactivated; (b) vector containing at least one
 XX gene structure of (a); (c) biotechnical production of organic compounds
 XX (particularly alcohols, aldehydes or organic acids) using (A); and (d)
 XX production method of (A). (A) are used to prepare organic compounds,
 XX preferably alcohols, aldehydes and organic acids, especially CA1, CA2,
 XX FA, and/or VA, most particularly a few known intermediates, e.g.,
 XX CA1, CA2, and/or VA, which are particularly useful for the production of
 XX single step. This sequence represents the Pseudomonas sp. HR199 vanillin
 XX dehydrogenase, vdh-omega-Km which is described in the method of the
 XX invention.
 XX
 XX Sequence 673 AA;
 XX
 XX Query Match 30.8%; Score 45; DB 21; Length 673;
 XX Best Local Similarity 47.1%; Pred. NO. 2.4e+02;
 XX Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 OY 4 GLYGCVPRAEPCTEARR 20
 DB 81 glysgvpkspvlgvqr 97
 XX
 XX RESULT 45
 XX AAB19929
 XX ID AAB19929 standard; Protein; 157 AA.
 XX
 XX AAB19929;
 XX
 XX 19-MAR-2001 (first entry)
 XX
 XX Human oxidoreductase OXRD-4.
 XX
 XX OXRD-4, human, oxidoreductase; cell proliferation;
 XX neurological disease; smooth muscle disease; autoimmune disease;
 XX inflammation; antiproliferative; neuroprotective;
 XX immunosuppressive; antiinflammatory; therapy; diagnosis.
 XX
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Modified-site /note= "O-phosphorylated"
 XX Modified-site 29
 XX Modified-site /note= "O-phosphorylated"
 XX Modified-site 105
 XX Domain /note= "O-phosphorylated"
 XX Domain 92
 XX /note= "nitrogenase Mo/Fe protein domain"
 XX
 XX WO200071679-A2.

PD 30-NOV-2000.
 XX
 XX 19-MAY-2000: 2000WO-US13879.
 XX
 XX 20-MAY-1999: 99US-0135049.
 XX
 XX 27-MAY-1999: 99US-0136740.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Yue H, Tang YT, Baughn MR, Lu DAM;
 XX
 XX WPI: 2001-025146/03.
 XX N-PSDB; AAB89188.
 XX
 XX New human oxidoreductase proteins useful for diagnosing, treating or
 XX preventing proliferative, neurological, genetic, smooth muscle,
 XX autoimmune or inflammatory disorders associated with abnormal
 XX expression of oxidoreductase proteins
 XX
 XX Claim 1(a): Page 84; 95pp; English.
 XX
 XX The present sequence is that of human oxidoreductase OXRD-4, as
 XX deduced from a cDNA clone (see AAB89188) isolated from a foetal
 XX kidney cDNA library. The protein shows homology to Bacillus subtilis
 XX nitrite reductase subunit, and includes a nitrogenase Mo/Fe protein
 XX domain. OXRD-4 is expressed in a prokaryotic host, and may be involved
 XX in cell proliferation and inflammatory tissue, and may be involved
 XX in cell proliferation and inflammation. The invention provides
 XX OXRD-1 to -8 polypeptides (see AAB19926-33) and polynucleotides (see
 XX AAB89185-92). It also provides methods for using these polypeptides
 XX and polynucleotides for diagnosing, treating or preventing disorders
 XX associated with expression of OXRD, especially cell proliferative,
 XX neurological, genetic, smooth muscle, autoimmune or inflammatory,
 XX somatic or germline gene therapy for treating such disorders.
 XX The proteins can also be used to screen for agonists and
 XX antagonists useful for treating these conditions, while antibodies
 XX that bind to OXRD may be used for diagnosis or in assays to monitor
 XX patient treatment.
 XX
 XX Sequence 157 AA;

Query Match 30.5%; Score 44.5; DB 22; Length 157;
 Best Local Similarity 41.4%; Pred. NO. 57;
 Matches 12; Conservative 3; Mismatches 5; Indels 9; Gaps 2;

OY 2 GEGLYGCVPRAEPCTEARRH 21
 DB 92 geglygsinpkpsakpkwckgikgrh 120

Search completed: June 28, 2001, 11:54:17
 Job time: 78 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:01 ; Search time 42.81 Seconds
(without alignments)
48.043 Million cell updates/sec

Title: US-09-439-313-566

Perfect score: 146

Sequence: 1 VGBGLYQGVPRAPGPGTEARRHYDEGVR 27

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database :

PIR_68:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	43.2	742	1 A43344	synaptic vesicle p
2	63	43.2	742	2 S27263	synaptic vesicle p
3	50.5	34.6	610	2 F83384	probable asparagin
4	49	33.6	420	2 T50585	probable membrane
5	49	33.6	2314	2 T28698	hypothetical prote
6	48	32.9	153	2 T37219	probable lipoprote
7	48	32.9	558	1 S03707	glutamate dehydrog
8	48	32.9	616	2 JQ1441	hypothetical 67K p
9	47.5	32.5	175	2 T09249	late embryonic abu
10	47.5	32.5	309	1 XNECV	branched-chain-ami
11	47.5	32.5	413	2 JC2520	beta-fructofuranos
12	47	32.2	127	2 A82594	Apag protein XF214
13	47	32.2	346	2 C75139	bifunctional short
14	47	32.2	457	1 A28468	chromogranin A pre
15	47	32.2	484	2 JE0341	R-SMAD protein, Sm
16	47	32.2	558	1 DEHDE	glutamate dehydrog
17	47	32.2	921	2 H84921	hypothetical prote
18	47	32.2	962	2 G86479	hypothetical prote
19	46.5	31.8	110	2 C75382	hypothetical prote
20	46.5	31.8	309	2 A86063	branched-chain ami
21	46	31.5	249	2 T09407	stationary-phase s
22	46	31.5	529	2 S68120	laccase [EC 1.10.3
23	46	31.5	558	1 A53719	glutamate dehydrog
24	46	31.5	583	2 B64816	ABC-type transport
25	46	31.5	583	1 A85588	hypothetical prote
26	46	31.5	592	2 E75032	carbon starvation
27	46	31.5	744	1 QOBEA7	U69 protein - hum
28	46	31.5	1741	2 T15978	hypothetical prote
29	45.5	31.2	105	2 A43861	vir-repressed prot

30	45.5	31.2	308	1 A34082	branched-chain-ami
31	45.5	31.2	413	2 S47527	extracellular suc
32	45.5	31.2	714	2 T00942	hypothetical prote
33	45.5	31.2	806	2 T01164	hypothetical prote
34	45.5	31.2	819	2 A84504	Mutator-like trans
35	45	30.8	38	2 B29502	carboxylesterase (
36	45	30.8	324	2 JC2395	Fas antigen precu
37	45	30.8	557	2 S38908	glutathione reduct
38	45	30.8	565	1 A65012	hypothetical prote
39	45	30.8	565	2 F85880	probable sensor pr
40	45	30.8	582	2 A84206	hypothetical prote
41	45	30.8	1460	1 EDBE1F	immediate-early pr
42	45	30.8	2124	2 A28452	proteoglycan core
43	44.5	30.5	142	2 E84396	hypothetical prote
44	44.5	30.5	235	2 E72783	hypothetical prote
45	44.5	30.5	413	2 E70661	probable PE protei
46	44	30.1	196	2 H72461	hypothetical prote
47	44	30.1	202	2 F72739	hypothetical prote
48	44	30.1	213	2 C82953	hypothetical prote
49	44	30.1	215	1 E72546	L-fucose-phospha
50	44	30.1	216	2 JC5039	adenylate kinase (
51	44	30.1	216	2 G84324	adenylate kinase (
52	44	30.1	217	1 K1FC	adenylate kinase (
53	44	30.1	220	2 B84304	hypothetical prote
54	44	30.1	262	2 E75478	conserved hypothet
55	44	30.1	371	2 E72665	hypothetical prote
56	44	30.1	406	2 T47674	hypothetical prote
57	44	30.1	562	2 JC5182	serine C-palmitoyl
58	44	30.1	595	2 T49384	related to NRD1 pr
59	44	30.1	730	2 T16455	hypothetical prote
60	44	30.1	809	1 C56279	carbon-monoxide de
61	44	30.1	926	2 H84415	DNA-directed RNA p
62	44	30.1	972	1 A33926	DNA-directed RNA p
63	44	30.1	1323	2 PNO568	connectin 3B - chl
64	44	30.1	7962	2 I38346	elastic titin - hu
65	43.5	29.8	152	2 JQ1244	coat protein 2 - b

ALIGNMENTS

RESULT	1
A43344	synaptic vesicle protein SV2 - rat
N:Alternate names: transporter-like protein p87	
C:Species: Rattus norvegicus (Norway rat)	
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000	
C:Accession: A43344; A43267; A58948	
R:Peany, M.B.; Lee, S.; Edwards, R.H.; Buckley, K.M.	
Cell 70, 861-867, 1992	
A>Title: The synaptic vesicle protein SV2 is a novel type of transmembrane transporte	
A:Reference number: A43344; MUID:92386605	
A:Accession: A43344	
A:Molecule type: mRNA	
A:Residues: 1-742 <FEA>	
A:Cross-references: GB:L01788; NID:g207090	
A:Experimental source: pheochromocytoma cell line PC12	
A>Note: sequence extracted from NCBI backbone (NCBIN:112840, NCBIP:112842)	
A>Note: this ORF is not annotated in GenBank entry RATSV2A, release 109.0	
R:Bajjalieh, S.M.; Peterson, K.; Shinghal, R.; Scheller, R.H.	
Science 257, 1271-1273, 1992	
A>Title: SV2, a brain synaptic vesicle protein homologous to bacterial transporters.	
A:Reference number: A43267; MUID:92390722	
A:Accession: A43267	
A:Molecule type: mRNA	
A:Residues: 1-339, 'F', 341-742 <BAJ>	
A:Cross-references: GB:L05435; NID:g207091; PIDN:AAA42188.1; PID:g207092	
A:Experimental source: brain	
A>Note: sequence extracted from NCBI backbone (NCBIN:112489)	
R:Gingrich, J.A.; Andersen, P.H.; Tiberi, M.; El Mestikawy, S.; Jorgensen, P.N.; From	
FEBS Lett. 312, 115-122, 1992	
A>Title: Identification, characterization, and molecular cloning of a novel transport	
A:Reference number: S27263; MUID:93050176	

[illegible]

A:Title: The complete genome sequence of *Escherichia coli* K-12.


```

Regul. Pept. 13, 207-233, 1986
A:Title: Chromogranin from normal human adrenal glands: purification by monoclonal an
A:Reference number: A23474; MUID:86206941
A:Accession: A23474
A:Molecule type: protein
A:Residues: 19-46 <WIL>
R:Seikiya, K.; Ghatei, M.A.; Minamino, N.; Bretherton-Watt, D.; Matsuo, H.; Bloom, S.R.
FEBS Lett. 228, 153-156, 1988
A:Title: Isolation of human pancreasastatin fragment containing the active sequence fro
A:Reference number: S00313; MUID:88137586
A:Accession: S00313
A:Molecule type: protein
A:Residues: 291-307/309-319 <SEK>
R:Curry, W.J.; Shaw, C.; Johnston, C.F.; Thim, L.; Buchanan, K.D.
FEBS Lett. 301, 319-321, 1992
A:Title: Isolation and primary structure of a novel chromogranin A-derived peptide, w
A:Reference number: S21141; MUID:92249591
A:Accession: S21141
A:Status: preliminary
A:Molecule type: protein
A:Residues: 342-355 <CUR>
C:Comment: Although human chromogranin A contains a number of paired basic residue si
tides, such peptides have not yet been characterized in the human.
C:Genetics:
A:Gene: GDB:CHGA
A:Cross-references: GDB:119777; OMIM:118910
A:Map position: 14q32-14q32
C:Superfamily: chromogranin A
C:Keywords: amidated carboxyl end; calcium binding; glycoprotein
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-457/Product: chromogranin A #status predicted <WAT>
F:219-307/309-319/Product: pancreasastatin-like peptide #status experimental <PLP>
F:278-287/Domain: calcium binding #status predicted <CAL>
F:35-56/Disulfide bonds: #status experimental
F:110/Binding site: carboxylate (Asn) (covalent) #status predicted
F:319/Modified site: amidated carboxyl end (Gly) (amide in mature form from following

```


C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: G86479
 R:Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: G86479
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-962 <STO>
 A:Cross-references: GB:AE005172; NID:g8778971; PIDN:AAF79886.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 32.2%; Score 47; DB 2; Length 962;
 Best Local Similarity 70.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VGEGLYQGV 10
 : : : : :
 DB 820 IGDLYEGVP 829

RESULT 19

C:Species: Deinococcus radiodurans (strain R1)
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
 C:Accession: C75382
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: C75382
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-110 <WHI>
 A:Cross-references: GB:AE001998; GB:AE000513; NID:g6459302; PIDN:AAF11112.1; PID:g645931
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR1539
 A:Map position: 1
 C:Superfamily: Deinococcus radiodurans hypothetical protein DR1539

Query Match 31.8%; Score 46.5; DB 2; Length 110;
 Best Local Similarity 46.4%; Pred. No. 13;
 Matches 13; Conservative 2; Mismatches 10; Indels 3; Gaps 1;

QY 3 EGYLQGVPAEPG---TEARRHYDEGVR 27
 | : | | | : | : | | | : |
 DB 75 EKVAGVDRAKAGIHNAEAHASYDEGHR 102

RESULT 20

A:Accession: A86063
 branched-chain amino-acid aminotransferase [imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: A86063
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: A86063
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-309 <STO>
 A:Cross-references: GB:AE005174; NID:g12518628; PIDN:AAG58965.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ilvE
 C:Superfamily: branched-chain-amino-acid aminotransferase

Query Match 31.8%; Score 46.5; DB 2; Length 309;
 Best Local Similarity 31.4%; Pred. No. 38;
 Matches 16; Conservative 3; Mismatches 5; Indels 27; Gaps 3;

QY 3 EGYLQGVPAEPG---TEARRHYDEGVR 26
 | : | | | : | : | | | : |
 DB 134 EALEQGDAMVSSWNRAPNTIPTAAKAGNYLSLLVGSSEARRHYQEG 184

RESULT 21

T09407
 stationary-phase survival protein sure - Haemophilus influenzae (strain Rd)
 C:Species: Haemophilus influenzae
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 23-Jul-1999
 C:Accession: T09407
 R:White, O.; Clayton, R.A.; Kerlavage, A.R.; Fleischmann, R.D.; Peterson, J.; Hickey,
 submitted to the EMBL Data Library, May 1998
 A:Reference number: Z16667
 A:Accession: T09407
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-249 <WHI>
 A:Cross-references: EMBL:U32753; NID:g1573701; PID:g3212202
 C:Genetics:
 A:Gene: HI0702
 C:Superfamily: stationary-phase survival protein Sure

Query Match 31.5%; Score 46; DB 2; Length 249;
 Best Local Similarity 32.0%; Pred. No. 36;
 Matches 8; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 3 EGYLQGVPAEPGTEARRHYDEGVR 27
 | : | : | : | : | : | : |
 DB 111 EGRHLGLPAITAVSLDGRQHYETAAR 135

RESULT 22

S68120
 laccase (EC 1.10.3.2) 4 precursor - Rhizoctonia solani
 C:Species: Rhizoctonia solani
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
 C:Accession: S68120; S77976; S72564
 R:Wahlthethner, J.A.; Xu, F.; Brown, K.M.; Brown, S.H.; Golightly, E.J.; Halkier, T.;
 Curr. Genet. 29, 395-403, 1996
 A:Title: The identification and characterization of four laccases from the plant path
 A:Reference number: S68117; MUID:96171523
 A:Accession: S68120
 A:Molecule type: DNA
 A:Residues: 1-529 <WAH1>
 A:Accession: S77976
 A:Molecule type: protein
 A:Residues: 20-53;132-180;203-217;378-396;440-465;489-513;520-529 <WAH2>
 R:Wahlthethner, J.A.
 submitted to the EMBL Data Library, September 1995
 A:Reference number: S72564
 A:Accession: S72564
 A:Molecule type: DNA
 A:Residues: 1-417; 'GAQHHR', 424-469, 'A', 471-483, 'CT', 484, 'R', 486, 'GSH', 490, 'S', 492, 'PR

A::Title: The complete genome sequence of Escherichia coli K-12.
A::Reference number: A64720; MUID:97426617
A::Accession: B64816
A::Status: nucleic acid sequence not shown; translation not shown
A::Molecule type: DNA
A::Residues: 1-583 <BLAT>
A::Cross-references: GB:AEO00181; GR:UO0096; NID:g1786998; PIDN:AAC73881.1; PID:g17870
A::Experimental source: strain K-12, substrain MG1655
A::Genetic info:
C::Superfamily: Escherichia coli ABC transporter ybhF; ATP-binding cassette homology
C::Keywords: Atyp; nucleotide binding; P-loop
F::28-218/Domain: ATP-binding cassette homology <ABC1>
F::45-52/Region: nucleotide-binding motif A (P-loop)
F::350-540/Domain: ATP-binding cassette homology <ABC2>
F::367-374/Region: nucleotide-binding motif A (P-loop)

Query Match 31.5%; Score 46; DB 1; Length 583;
Best Local Similarity 52.4%; Pred. No. 87;
Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 GEGLYGVVPRAEPTGTEARRHY 22
DB 218 GELLTQGEKALQTWAGRSF 238

RESULT 25

A85588
hypothetical protein ybhF [imported] - Escherichia coli (strain O157:H7)
C::Species: Escherichia coli
C::Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
R::Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May, M.A.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001
A::Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A::Reference number: A85480; MUID:21074935; PMID:11206551
A::Accession: AF558694
A::Status: Genomic
A::Molecule type: DNA
A::Residues: 1-583 <STO>
A::Cross-references: GB:AEO05174; NID:g12513786; PIDN:AAG55165.1; GSPDB:GN00145; UWGP:
A::Experimental source: strain O157:H7, substrain EDL933
C::Genetics:
A::Gene: ybhF
C::Superfamily: Escherichia coli ABC transporter ybhF; ATP-binding cassette homology

Query Match 31.5%; Score 46; DB 2; Length 583;
Best Local Similarity 52.4%; Pred. No. 87;
Matches 11; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 GEGLYGVVPRAEPTGTEARRHY 22
DB 218 GELLTQGEKALQTWAGRSF 238

RESULT 26

E75032
Carbon starvation protein A homolog PAB1554 [similarity] - Pyrococcus abyssi (strain ATCC 49239) [similarity]
C::Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 18-Aug-2000
C::Accession: E75032
R::anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A::Description: Pyrococcus abyssi genome sequence
A::Reference number: A75001
A::Accession: E75032
A::Status: Preliminary
A::Molecule type: DNA
A::Residues: 1-592 <KAW>
A::Cross-references: GB:AJ248287; GB:AL096836; NID:g5458657; PIDN:CAB50154.1; PID:g54545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB1554

C:Superfamily: carbon starvation protein

Query Match 31.5%; Score 46; DB 2; Length 592;
Best Local Similarity 50.0%; Pred. No. 89;
Matches 12; Conservative 2; Mismatches 8; Indels 2; Gaps 1;

QY 2 GEGLYQGVPRAPG--TEARRHYD 23

Db 22 GKGLQNKVVVRADPNRPTPAHRLYD 45

RESULT 27

QBEA7

UL69 protein - human cytomegalovirus (strain AD169)

C:Species: human cytomegalovirus, human herpesvirus 5

A:Note: host Homo sapiens (man)

C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999

C:Accession: S09832

R:Chen, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.; M.; Barrall, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A:Reference number: S09749; MUID:90269039

A:Accession: S09832

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-744 <CHE>

A:Cross-references: EMBL:X17403; NID:g59591; PIDN:CRA35384.1; PID:gl780846

A:Note: possible protein-coding frames are given

A:Note: The DNA sequence was submitted to EMBL, December 1989, in computer-readable form

C:Superfamily: cytomegalovirus UL69 protein

C:Keywords: transcription regulation

Query Match 31.5%; Score 46; DB 1; Length 744;
Best Local Similarity 43.8%; Pred. No. 11e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 10 PRAEPTGTEARRHYDEG 25

Db 598 PPAQPPSQPOQHSEG 613

RESULT 28

T15978

hypothetical protein F08F8.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 23-Mar-2001

C:Accession: T15978

R:Du, Z.

submitted to The EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid F08F8.

A:Reference number: S61147

A:Accession: T15978

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1741 <DUZ>

A:Cross-references: EMBL:U28991; NID:g861364; PID:g861366; PIDN:AAA68385.1; CESP:F08F8.4

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F08F8.4

A:Introns: 33/1; 64/1; 130/3; 201/2; 251/3; 362/2; 568/2; 628/3; 648/3; 655/3; 669/3; 737/2; 1676/3

Query Match 31.5%; Score 46; DB 2; Length 1741;
Best Local Similarity 40.9%; Pred. No. 2.7e+02;
Matches 9; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAPGTEARRHYD 23

Db 890 GSGYQHIGNTWPNTEQWNYD 911

RESULT 29

A43861

vir-repressed protein virg-6 - Bordetella pertussis

C:Species: Bordetella pertussis

C:Date: 10-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 18-Nov-1994

C:Accession: A43861

R:Beattie, D.T.; Shahin, R.; Mekalanos, J.J.

Infect. Immun. 60, 571-577, 1992

A:Title: A vir-repressed gene of Bordetella pertussis is required for virulence.

A:Reference number: A43861; MUID:92112321

A:Accession: A43861

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-105 <BEA>

A:Note: sequence inconsistent with the nucleotide translation

A:Note: sequence extracted from NCBI backbone (NCBIN:77525, NCBI:P:77526)

Query Match 31.2%; Score 45.5; DB 2; Length 105;
Best Local Similarity 47.8%; Pred. No. 17;
Matches 11; Conservative 1; Mismatches 8; Indels 3; Gaps 1;

QY 5 LYQGVPRAPGTEARRHYDEGVR 27

Db 74 VYGGPRYKGG---HRHYDRGRH 93

RESULT 30

A34082

branched-chain-amino-acid transaminase (EC 2.6.1.42) - Salmonella typhimurium

N:Alternate names: branched-chain-amino-acid aminotransferase; transaminase B

C:Species: Salmonella typhimurium

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999

C:Accession: A34082; A05077

R:Feild, M.J.; Nguyen, D.C.; Armstrong, F.B.

Biochemistry 28, 5306-5310, 1989

A:Title: Amino acid sequence of Salmonella typhimurium branched-chain amino acid amin

A:Reference number: A34082; MUID:89352621

A:Accession: A34082

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-308 <FEI>

R:Randall, R.R.; Wallis, M.H.; Young, G.J.; Armstrong, F.B.

Fed. Proc. 38, 325A, 1979

A:Title: N-terminal sequence of branched-chain amino acid aminotransferase.

A:Reference number: A05077

A:Accession: A05077

A:Molecule type: protein

A:Residues: 1-5 <RAN>

C:Function:

A:Description: catalyzes the reversible transamination of alpha-amino groups from leu

A:Pathway: branched-chain amino acid biosynthesis

C:Superfamily: branched-chain-amino-acid aminotransferase

C:Keywords: aminotransferase; branched-chain amino acid biosynthesis; phosphoprotein;

F:159/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 31.2%; Score 45.5; DB 1; Length 308;
Best Local Similarity 31.4%; Pred. No. 53;
Matches 16; Conservative 3; Mismatches 5; Indels 27; Gaps 3;

QY 3 EGLYQGV-----RAEP-----GTEARRH-YDEGV 26

Db 133 EALDQGDAMVSSNNRAAPNTIPTAAKAGGNLYSLLVGSEARRHGYOEGI 183

RESULT 31

S47527

C:Genetics:
A:Map position: 4
A:Introns: 380/3; 675/2
A>Note: F7N22.10
C:Superfamily: Arabidopsis thaliana hypothetical protein F26C24.7

Query Match 31.2%; Score 45.5; DB 2; Length 806;
Best Local Similarity 52.2%; Pred. NO. 1.4e+02;
Matches 12; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

QY 5 LYQGVPRAPGPTGARRHYDEGVR 27
|| : || ||| || ||| ||
DB 412 LYM-IRRANFGTVARLQIDESGR 433

RESULT 34
AB4504
Mutator-like transposase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Mar-2001
E:Accession: AB4504
E:Accession: K84534; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, S.; Doo, J.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, D.; Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487
A:Accession: AB4504
A:Status: Preliminary
A:Scatus: Preliminary
A:Residues: 1-819 <STO>
A:Cross-references: GB:AE002093; NID:g4850412; PIDN:AAD31079.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g12720
A:Map position: 2
C:Superfamily: Arabidopsis thaliana hypothetical protein F26C24.7

Query Match 31.2%; Score 45.5; DB 2; Length 819;
Best Local Similarity 52.2%; Pred. NO. 1.5e+02;
Matches 12; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

QY 5 LYQGVPRAPGPTGARRHYDEGVR 27
|| : || ||| || ||| ||
DB 407 LYM-IRRANFGTVARLQIDESGR 428

RESULT 35
B29502
Carboxylesterase (EC 3.1.1.1) 5 - fruit fly (Drosophila mojavensis) (fragment)
N:Alternate names: esterase-5
C:Species: Drosophila mojavensis
C:Date: 21-May-2002 #sequence_revision 21-May-1988 #text_change 11-Jan-2000
E:Accession: B29502
E:Accession: Van Beuningen, J.; Beintema, J.J.
Biochem. J. 278, 691-699, 1986
A:Title: Structural comparison of two esterases from Drosophila mojavensis isolated b
A:Reference number: A90340
A:Accession: B29502
A:Molecule type: protein
C:Residues: 1-58 <PEN>
A:Gene: FlyBase:Dmoj/Est-5
A:Cross-references: FlyBase:FBgn0012569
C:Superfamily: cholinesterase; cholinesterase homology
C:Keywords: carboxylic ester hydrolase

Query Match 30.8%; Score 45; DB 2; Length 38;
Best Local Similarity 64.3%; Pred. NO. 7.1;
Matches 9; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

QY 3 EGLY--QGVPRAP 14
 ||| :|||
 Db 22 EGYEAGIPRAEP 35

RESULT 36

JC2395
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-Nov-1999
 C:Accession: JC2395; PC2246
 R:Kimura, K.; Wakatsuki, T.; Yamamoto, M.
 Biochem. Biophys. Res. Commun. 198, 666-674, 1994
 A:Title: A variant mRNA species encoding a truncated form of Fas antigen in the rat liver
 A:Reference number: JC2395; MUID:94128114
 A:Accession: JC2395
 A:Molecule type: mRNA
 A:Residues: 1-324 <KIM>
 A:Cross-references: DBJ:D26112; NID:g468486; PIDN:BAA05108.1; PID:dl005650; PID:g468487
 A:Experimental source: thymus
 A:Accession: PC2246
 A:Molecule type: mRNA
 A:Residues: 1-62, RFT, <K12>
 A:Cross-references: DBJ:D26113; NID:g468488; PIDN:BAA05109.1; PID:dl005651; PID:g468489
 A:Experimental source: liver
 C:Genetics:
 A:Introns: 62/1
 C:Superfamily: NGF receptor repeat homology
 C:Keywords: transmembrane protein
 F:1-21/Domain: signal sequence predicted <SIG>
 F:22-324/Product: Fas antigen #status predicted <MAT>
 F:44-79/Domain: NGF receptor repeat homology <NGF>
 F:81-124/Domain: NGF receptor repeat homology <NG4>
 F:171-186/Domain: transmembrane #status predicted <TMM>

Query Match 30.8%; Score 45; DB 2; Length 324;
 Best Local Similarity 26.4%; Pred. No. 66;
 Matches 14; Conservative 2; Mismatches 9; Indels 28; Gaps 2;

QY 3 EGLYQ-----GVPRAPGTEA-----RRHYDEGVR 27
 ||||| :|||
 Db 46 EGLYQVGPPCCQPCQGERKVKDCITSGAPTCHPCTEGEYTRKHYSDKCR 98

RESULT 37

S38908
 glutathione reductase (NADPH) (EC 1.6.4.2) - common tobacco (fragment)
 C:Species: Nicotiana tabacum (common tobacco)
 C:Date: 22-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 03-Nov-2000
 C:Accession: S38908
 R:Creissen, G.; Mullineaux, X.Y.Z.
 submitted to the EMBL Data Library, November 1993
 A:Reference number: S38908
 A:Accession: S38908
 A:Molecule type: mRNA
 A:Residues: 1-557 <CRE>
 A:Cross-references: EMBL:X76293; NID:g431954; PIDN:CAA53925.1; PID:g431955
 C:Superfamily: dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase homology
 C:Keywords: FAD; flavoprotein; NADP; oxidoreductase; redox-active disulfide
 F:83-528/Domain: dihydrolipoamide dehydrogenase homology <BLD>
 F:127-132/Disulfide bonds: redox-active #status predicted

Query Match 30.8%; Score 45; DB 2; Length 557;
 Best Local Similarity 62.5%; Pred. No. 1.2e+02;
 Matches 10; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 10 PRAEP--GTEARRHYD 23
 ||||| :|||
 Db 64 PRAESSNGADAPRHYD 79

RESULT 38

A65012
 hypothetical protein b2380 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: A65012
 R:Blattner, F.R.; Mau, B.; Shao, Y.
 A.; Rose, D.J.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: A65012
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-565 <BLAT>
 A:Cross-references: GB:AE000326; GB:U00096; NID:g1788718; PIDN:AAC75439.1; PID:g17887
 A:Experimental source: strain K-12, substrain MG1655
 C:Superfamily: hypothetical protein b2380

Query Match 30.8%; Score 45; DB 1; Length 565;
 Best Local Similarity 62.5%; Pred. No. 1.2e+02;
 Matches 10; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 GEGLYQGVPRAPGTE 17
 ||||| :|||
 Db 532 GEGLH--IRRLPGE 545

RESULT 39

F85880
 probable sensor protein Z3645 [imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: F85880
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamasis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: F85880
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-565 <STO>
 A:Cross-references: GB:AE005174; NID:g12516754; PIDN:AAG57506.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z3645
 C:Superfamily: hypothetical protein b2380

Query Match 30.8%; Score 45; DB 2; Length 565;
 Best Local Similarity 62.5%; Pred. No. 1.2e+02;
 Matches 10; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 GEGLYQGVPRAPGTE 17
 ||||| :|||
 Db 532 GEGLH--IRRLPGE 545

RESULT 40

A84206
 hypothetical protein porA [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: A84206
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
 ; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483

A:Title: Cartilage proteoglycan aggregates. The link protein and proteoglycan amino-
A:Reference number: A28453; MUID:88087071
A:Accession: A28453
A:Molecule type: protein
A:Residues: 20-37; W:39-60; E:62-64; X':66-69;70-83;84-89-148;'L':150-238;'S':240;
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: glycoprotein
E:20-3124;Protein sequence status predicted <SIG>
F:20-2124;Protein sequence status predicted <SIG>
F:44-135;Domain: immunoglobulin homology <IMM>
F:170-247;Domain: link protein repeat homology <LNK1>
F:268-349;Domain: link protein repeat homology <LNK2>
F:504-581;Domain: link protein repeat homology <LNK3>
F:602-683;Domain: link protein repeat homology <LNK4>
F:683-707;Domain: link protein repeat homology <LNK5>
F:2043-2097;Domain: complement factor H repeat homology <PHD>
F:126-239,333,387,611,667,1842;Binding site: carbohydrate (Asn) (covalent) #status pr
F:126,239,333,387,611,667,1842;Binding site: carbohydrate (Asn) (covalent) #status pr

Query Match 30.8%; Score 45; DB 2; Length 2124;
Best Local Similarity 43.5%; Pred. No: 4.7e+02;
Matches 10; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

QY 4 GLYQGVPRAPCGTEARRHYDEGV 26
Db 1100 GYVSGIPSGEDGTSTSGVEGV 1122
|||||
|||

RESULT 43
E84396
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84396
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky,
Jung, K.H.; Alam, M.; Freitas, T. 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: E84396
A:Status: preliminary
A:Molecule type: DNA
A:Cross-references: GB:AE004437; NID:g10581861; PIDN:AAG20537.1; GSPDB:GN00138
C:Genetics:
A:Gene: E84396

Query Match 30.5%; Score 44.5; DB 2; Length 142;
Best Local Similarity 52.2%; Pred. No: 3;
Matches 12; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 5 LYQGVPRAPCGTEARRHYDEGVR 27
Db 105 LLVGVPWPFG-EVRAHNRHGR 126
|||||
|||

RESULT 44
E72783
hypothetical protein APE0254 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: E72783
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Tanaka,
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
DNA Res. 6, 83-91, 1999
A:Reference number: A72450; MUID:99310339
A:Accession: E72783
A:Status: preliminary
A:Molecule type: DNA

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:12 ; Search time 37.02 Seconds
(without alignments)
14.692 Million cell updates/sec

Title: US-09-439-313-566
Perfect score: 146
Sequence: 1 VGEGLQGVRAEPGTEARRHYDEGVR 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues
Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	89.7	27	4	US-09-071-710-39
2	131	89.7	255	4	US-09-071-710-36
3	46.5	31.8	140	1	US-08-664-596B-2
4	46.5	31.8	140	2	US-08-738-367-2
5	46.5	31.8	240	2	US-08-738-367-7
6	46	31.5	529	1	US-08-172-331B-14
7	46	31.5	529	1	US-08-706-037-25
8	46	31.5	529	2	US-09-005-397-25
9	46	31.5	529	2	US-09-032-315-5
10	46	31.5	529	2	US-08-993-318A-5
11	46	31.5	529	4	US-09-399-886-5
12	46	31.5	529	4	US-09-396-260-5
13	46	31.5	564	2	US-08-865-337A-1
14	46	31.5	610	2	US-08-865-337A-6
15	43.5	29.8	4544	1	US-08-469-486-52
16	43.5	29.8	4544	2	US-08-469-658-52
17	43	29.5	113	2	US-08-563-892A-30
18	43	29.5	323	2	US-08-435-149-2
19	43	29.5	324	2	US-08-310-416A-14
20	43	29.5	324	2	US-08-888-171-14
21	43	29.5	577	2	US-08-435-149-3
22	42.5	29.1	445	1	US-08-674-168-30
23	42.5	29.1	445	3	US-08-985-908-20
24	42.5	29.1	445	3	US-08-852-730-5
25	42.5	29.1	445	4	US-08-985-916-12
26	42	28.8	348	3	US-08-652-265-4
27	42	28.8	348	3	US-08-652-265-8

Sequence 4, Appli
Sequence 8, Appli
Sequence 4, Appli
Sequence 8, Appli
Sequence 8, Appli
Sequence 8, Appli
Sequence 8, Appli
Sequence 9, Appli
Sequence 7, Appli
Sequence 8, Appli
Sequence 23, Appli
Sequence 2, Appli
Sequence 10, Appli
Sequence 29, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 3, Appli
Sequence 3, Appli
Sequence 106, App
Sequence 2, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 1, Appli
Sequence 66, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 12, Appli
Sequence 5, Appli
Sequence 4, Appli
Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-071-710-39
; Sequence 39, Application US/09071710
; Patent No. 6130043
GENERAL INFORMATION:
APPLICANT: BILLING-MEDEL, PATRICIA
APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GORDON, JULIAN
APPLICANT: GRANADOS, EDWARD N.
APPLICANT: HODGES, STEVEN C.
APPLICANT: KLASS, MICHAEL R.
APPLICANT: KRATOCHVIL, JON D.
APPLICANT: ROBERTS-RAPP, LISA
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STROUPE, STEPHEN D.
TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
FOR DETECTING DISEASES OF THE PROSTATE
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

Query Match	89.7%	Score 131;	DB 4;	Length 255;
Best Local Similarity	100.0%	Pred. No. 5.4e-12;		
Matches 24.	Conservative	Mismatches 0;	Indels 0;	Gaps 0;

```
Query Match      89.7%;   Score 131;   DB 4;   Length 255;
Best Local Similarity 100.0%;   Pred. No. 5.4e-12;
Matches 24;   Conservative 0;   Mismatches 0;   Indels
```

QY 4 GLYQGVPR AEPGTEARRHYDEGVR 27
DB 1 GLYQGVPR AEPGTEARRHYDEGVR 24

RESULT 3
US-08-664-596B-2
; Sequence 2, Application US/08664596B
; Patent No. 5807703

GENERAL INFORMATION: Kenneth
APPLICANT: Jacobs, Jr.
APPLICANT: Joseph, Albert
APPLICANT: Lavallee, Edward
APPLICANT: Racie, Lisa
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Swans, Cheryl
APPLICANT: Swans, Cheryl Vikki
APPLICANT: Bowman, Michael
APPLICANT: Bowman, Michael
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 37
NUMBER OF SEQUENCES: 37
CONTRACT NUMBER: 37
CONTRACT NUMBER: 37
ADDRESS: Genetics Institute, Inc.
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
STREET: 87 Cambridgepark Drive
CITY: Cambridge
CITY: Cambridge
STATE: Massachusetts
STATE: Massachusetts
COUNTRY: U. S. A.
COUNTRY: U. S. A.
MAILING ADDRESS: 314
MAILING ADDRESS: 314
COMPUTER READABLE FORM:
COMPUTER TYPE: Floppy disk
COMPUTER TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: ease #1.0, Version #1.30
CURRENT APPLICATION DATA: ease #1.0, Version #1.30
APPLICATION NUMBER: US/08/664, 596B
APPLICATION NUMBER: US/08/664, 596B

FILING DATE: 514
 CLASSIFICATION: SECRET INFORMATION:
 NAME: BROWN SCOTT A.
 REGISTRATION NUMBER: 32,724
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 875-5832
 TELETYPE: (617) 875-5832
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 140 amino acids
 TYPE: amino acid
 STRANDEDNESS:

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-664-596B-2

Query Match 31.8%; Score 46.5; DB 1; Length 140;
Best Local Similarity 47.4%; Pred. No. 8.4;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYOGVPRAPGTEARRHYD 23
:: ||| ||| | | |
Db 83 VWSGVPAEMPGRK-RHHFD 100

RESULT 4

US-08-738-367-2
; Sequence 2, Application US/08738367
; Patent No. 5827688

; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,367
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-738-367-7

Query Match 31.8%; Score 46.5; DB 2; Length 140;
Best Local Similarity 47.4%; Pred. No. 8.4;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYOGVPRAPGTEARRHYD 23
:: ||| ||| | | |
Db 83 VWSGVPAEMPGRK-RHHFD 100

RESULT 5

US-08-738-367-7
; Sequence 7, Application US/08738367

; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,367
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-738-367-2

Query Match 31.8%; Score 46.5; DB 2; Length 140;
Best Local Similarity 47.4%; Pred. No. 8.4;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYOGVPRAPGTEARRHYD 23
:: ||| ||| | | |
Db 83 VWSGVPAEMPGRK-RHHFD 100

RESULT 6

US-08-172-331B-14
; Sequence 14, Application US/08172331B
; Patent No. 5480801

; GENERAL INFORMATION:
; APPLICANT: Wahleithner, Jill A.
; APPLICANT: Christensen, Bjoern E.
; APPLICANT: Schneider, Pallo
; TITLE OF INVENTION: PURIFIED PH NEUTRAL LACCASES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5480801disk of No. 5480801th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
US-08-172-331B-14

Query Match 31.8%; Score 46.5; DB 2; Length 140;
Best Local Similarity 47.4%; Pred. No. 8.4;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYOGVPRAPGTEARRHYD 23
:: ||| ||| | | |
Db 83 VWSGVPAEMPGRK-RHHFD 100

RESULT 7

US-08-738-367-7
; Sequence 7, Application US/08738367

; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
US-08-738-367-2

; Patent No. 5827688
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Lavallie, Edward
; APPLICANT: Racie, Lisa
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,367
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-738-367-7

Query Match 31.8%; Score 46.5; DB 2; Length 240;
Best Local Similarity 47.4%; Pred. No. 16;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYOGVPRAPGTEARRHYD 23
:: ||| ||| | | |
Db 83 VWSGVPAEMPGRK-RHHFD 100

RESULT 8

US-08-172-331B-14
; Sequence 14, Application US/08172331B
; Patent No. 5480801

; GENERAL INFORMATION:
; APPLICANT: Wahleithner, Jill A.
; APPLICANT: Christensen, Bjoern E.
; APPLICANT: Schneider, Pallo
; TITLE OF INVENTION: PURIFIED PH NEUTRAL LACCASES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5480801disk of No. 5480801th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
US-08-172-331B-14

Query Match 31.8%; Score 46.5; DB 2; Length 240;
Best Local Similarity 47.4%; Pred. No. 16;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYOGVPRAPGTEARRHYD 23
:: ||| ||| | | |
Db 83 VWSGVPAEMPGRK-RHHFD 100

RESULT 9

US-08-172-331B-14
; Sequence 14, Application US/08172331B
; Patent No. 5480801

; GENERAL INFORMATION:
; APPLICANT: Wahleithner, Jill A.
; APPLICANT: Christensen, Bjoern E.
; APPLICANT: Schneider, Pallo
; TITLE OF INVENTION: PURIFIED PH NEUTRAL LACCASES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5480801disk of No. 5480801th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
US-08-172-331B-14

Query Match 31.8%; Score 46.5; DB 2; Length 240;
Best Local Similarity 47.4%; Pred. No. 16;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 5 LYOGVPRAPGTEARRHYD 23
:: ||| ||| | | |
Db 83 VWSGVPAEMPGRK-RHHFD 100

RESULT 10

US-08-172-331B-14
; Sequence 14, Application US/08172331B
; Patent No. 5480801

; GENERAL INFORMATION:
; APPLICANT: Wahleithner, Jill A.
; APPLICANT: Christensen, Bjoern E.
; APPLICANT: Schneider, Pallo
; TITLE OF INVENTION: PURIFIED PH NEUTRAL LACCASES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5480801disk of No. 5480801th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
US-08-172-331B-14

Query Match 31.8%; Score 46.5; DB 2; Length 240;
Best Local Similarity 47.4%; Pred. No. 16;
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/172.331B
FILING DATE: 22-DEC-1993
CLASSIFICATION: 435
PRIORITY INFORMATION DATA:
PRIORITY NUMBER: US 08/122.230
FILING DATE: 17-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/122.827
FILING DATE: 17-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/162.827
FILING DATE: 03-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Lowney Dr., Karen A.
REGISTRATION NUMBER: 31,274
REFERENCE/DOCKET NUMBER: 4052.020-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-878-9655
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-172-331B-14

Query Match 31.5%; Score 46; DB 1; Length 529;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YOGVPRAPGTE 17

DB 314 YEGAPNAEPTTE 325

RESULT 706/037-25
US-08-172-331B-14
Sequence 25 Application US/08706037
Patent No. 5770419
GENERAL INFORMATION:
APPLICANT: Xu, Feng
APPLICANT: Berka, Randy M.
APPLICANT: Wahleithner, Jill A.
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH
ENHANCED ACTIVITY
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESS: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,037
FILING DATE: 30-AUG-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,800
FILING DATE: 30-AUG-1996
ATTESTATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4526.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-09-005-397-25

Query Match 31.5%; Score 46; DB 2; Length 529;
Best Local Similarity 66.7%; Pred. No. 46;

REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4526.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-706-037-25

Query Match 31.5%; Score 46; DB 1; Length 529;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YOGVPRAPGTE 17

DB 314 YEGAPNAEPTTE 325

RESULT 8
US-09-005-397-25
Sequence 25 Application US/09005397
Patent No. 59726700
GENERAL INFORMATION:
APPLICANT: Xu, Feng
APPLICANT: Berka, Randy M.
APPLICANT: Wahleithner, Jill A.
TITLE OF INVENTION: BLUE COPPER OXIDASE MUTANTS WITH
ENHANCED ACTIVITY
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESS: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,397
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/706,037
FILING DATE: 30-AUG-1996
APPLICATION NUMBER: US 60/002,800
FILING DATE: 30-AUG-1996
ATTESTATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4526.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-005-397-25

Query Match 31.5%; Score 46; DB 2; Length 529;
Best Local Similarity 66.7%; Pred. No. 46;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17
1:1111111111

Db 314 YEGAPNAEPTTE 325

RESULT 9
US-09-032-315-5
; Sequence 5, Application US/09032315
; Patent No. 5985818
; GENERAL INFORMATION:
; APPLICANT: Svendsen, Allan
; APPLICANT: Xu, Feng
; TITLE OF INVENTION: LACCASE MUTANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5985818o No. 5985818disk of No. 5985818th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/032.315
; FILING DATE: 27-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 5200.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-032-315-5

Query Match 31.5%; Score 46; DB 2; Length 529;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17
1:1111111111

Db 314 YEGAPNAEPTTE 325

RESULT 10
US-08-993-318A-5
; Sequence 5, Application US/08993318A
; Patent No. 5998353
; GENERAL INFORMATION:
; APPLICANT: Pedersen, Anders
; APPLICANT: Svendsen, Allan
; APPLICANT: Schneider, Palle
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Cherry, Joel
; TITLE OF INVENTION: LACCASE MUTANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5998353o No. 5998353disk of No. 5998353th America
; STREET: 405 Lexington Avenue

CITY: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993.318A
FILING DATE: December 18, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-993-318A-5

Query Match 31.5%; Score 46; DB 2; Length 529;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGTE 17
1:1111111111

Db 314 YEGAPNAEPTTE 325

RESULT 11
US-09-399-886-5
; Sequence 5, Application US/09399886
; Patent No. 6140092
; GENERAL INFORMATION:
; APPLICANT: Pedersen, Anders
; APPLICANT: Svendsen, Allan
; APPLICANT: Schneider, Palle
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Cherry, Joel
; TITLE OF INVENTION: LACCASE MUTANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6140092o No. 6140092disk of No. 6140092th America
; STREET: 405 Lexington Avenue
; CITY: New York
; COUNTRY: USA
; ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/399.886
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/993.318
FILING DATE: December 18, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-399-886-5

Query Match 31.5%; Score 46; DB 4; Length 529;
Best Local Similarity 66.7%; Pred No. 46;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 6 YQGVPRAPGTE 17
Db 314 YEGAPNAEPTTE 325

RESULT 12
US-09-396-260-5
Sequence 5, Application US/09396260
Patent No. 6184015
GENERAL INFORMATION:
APPLICANT: Xue Feng, Allan
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: NO. 61840150 No. 6184015disk of No. 6184015th America, Inc.
CITY: 405 Lexington Avenue
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
SOFTWARE: FASTSEQ for Windows Version 2.0
APPLICATION NUMBER: US/09/396,260
FILING DATE:
CLASSIFICATION: DATA:
PRIORITY/DOCKET NUMBER: US/09/032,315
FILING DATE: 27-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Rozek, Carol
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 5200.200-US
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-396-260-5

Query Match 31.5%; Score 46; DB 4; Length 529;
Best Local Similarity 66.7%; Pred No. 46;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 6 YQGVPRAPGTE 17

Db 314 YEGAPNAEPTTE 325

RESULT 13
US-08-865-337A-1
Sequence 1, Application US/08865337A
Patent No. 5972649
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Tang, Y. Tom
APPLICANT: Murty, Lynn E.
TITLE OF INVENTION: HUMAN MULTIPLE ENDOCRINE NEOPLASIA TYPE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/865,337A
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0305 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-853-0535
TELEFAX: 415-843-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
TYPE: amino acid
STRANDEDNESS: single
IMMEDIATE SOURCE:
LIBRARY: BRAINOT14
CLONE: Consensus
US-08-865-337A-1

Query Match 31.5%; Score 46; DB 2; Length 564;
Best Local Similarity 52.9%; Pred No. 49;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 2 GEGLYQGVPRAPGTEA 18
Db 461 GQGVSGPPRPPGIVA 477

RESULT 14
US-08-865-337A-6
Sequence 6, Application US/08865337A
Patent No. 5972649
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Tang, Y. Tom
APPLICANT: Murty, Lynn E.

;; TITLE OF INVENTION: HUMAN MULTIPLE ENDOCRINE NEOPLASTIA TYPE
;; TITLE OF INVENTION: 1 PROTEIN
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Incyte Pharmaceuticals, Inc.
;; STREET: 3174 Porter Drive
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94304
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/865,337A
;; FILING DATE: Herewith
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Billings, Lucy J.
;; REGISTRATION NUMBER: 36,749
;; REFERENCE/DOCKET NUMBER: PF-0305 US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-855-0555
;; TELEFAX: 415-845-4166
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 610 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; LIBRARY: GenBank
;; CLONE: 1945387
;; US-08-865-337A-6

Query Match 31.5%; Score 46; DB 2; Length 610;
Best Local Similarity 52.9%; Pred. No. 54;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 2 GEGYQGVPRAPGTEA 18
I:| | | | | | | |
Db 507 GQAVSGPPKPPGTVA 523

RESULT 15
US-08-469-486-52
; Sequence 52, Application US/08469486
; Patent No. 5739281
; GENERAL INFORMATION:
; APPLICANT: Thøgersen, Hans Christian
; APPLICANT: Holtet, Thor Las
; APPLICANT: Etzerodt, Michael
; TITLE OF INVENTION: Improved method for the refolding of
; TITLE OF INVENTION: proteins
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version
;; SOFTWARE: #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/469,486
;; FILING DATE:
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/192,060
;; FILING DATE: February 4, 1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Paul T. Clark
;; REGISTRATION NUMBER: 30,162
;; REFERENCE/DOCKET NUMBER: 06363/002001
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 617 542 5070
;; TELEFAX: 617 542 8906
;; TELEX: 200154
;; INFORMATION FOR SEQ ID NO: 52:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4544 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-469-486-52

Query Match 29.8%; Score 43.5; DB 1; Length 4544;
Best Local Similarity 41.7%; Pred. No. 1.3e+03;
Matches 10; Conservative 3; Mismatches 8; Indels 3; Gaps 1;

QY 6 YGVPRAPGTEARRH--YDEGV 26
I:| | | | | | | |
Db 3926 YRSLPPAAPTNSRRRQIDRGV 3949

RESULT 16
US-08-469-658-52
; Sequence 52, Application US/08469658
; Patent No. 5917018
; GENERAL INFORMATION:
; APPLICANT: Thøgersen, Hans Christian
; APPLICANT: Holtet, Thor Las
; APPLICANT: Etzerodt, Michael
; TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOLDING OF
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,658
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/192,060
; FILING DATE: February 4, 1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06363/002002
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 617 542 5070
TELEFAX: 617 542 8906
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 4544 amino acids
TYPE: amino acid
STRANDEDNESS:
MOLECULE TYPE: protein
US-08-469-658-52

Query Match 29.8%; Score 43.5; DB 2; Length 4544;
Best Local Similarity 41.7%; Pred. No. 1.3e+03;
Matches 10; Conservative 3; Mismatches 8; Indels 3; Gaps 1;

QY 6 YGVPRAPRGTEARRH---YDSGV 26
DB 3926 YRSLPPAAPPTTSNRHRQIDRGV 3949

RESULT 17
US-08-563-892A-30
Sequence 30, Application US/08563892A
Patent No. 5976819
GENERAL INFORMATION:
APPLICANT: Finkel, Terri H.
ATTORNEY/AGENT INFORMATION: Moshe M.
TITLE OF INVENTION: POLYMERIZATION IN T LYMPHOCYTES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
SOFTWARE: GENESYS/PC/MS-DOS
CURRENT APPLICATION DATA:
PATENT IN RELEASE #1.0, Version #1.30
APPLICATION NUMBER: US/08/563,892A
FILING DATE: 21-NOV-1995

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-33
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 113 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-563-892A-30

Query Match 29.5%; Score 43; DB 2; Length 113;
Best Local Similarity 47.4%; Pred. No. 22;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 3 EGLYQGVPRAPRGTEARRH 21
DB 88 DGLYQGLSTATKTDALH 106

RESULT 18
US-08-435-149-2
Sequence 2, Application US/08435149
Patent No. 5866402
GENERAL INFORMATION:
APPLICANT: INNIS, MICHAEL A.
ATTORNEY/AGENT INFORMATION: ZAROR, ISABEL A.
TITLE OF INVENTION: CHIMERIC MCP AND DAF PROTEINS WITH CELL
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: INTELLECTUAL PROPERTY - R440, P.O. BOX 8097
CITY: EMERYVILLE
STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94662-8097
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: XENIX PC compatible
SOFTWARE: SYMBIOSIS/MS-DOS
CURRENT APPLICATION DATA:
PATENT IN RELEASE #1.0, Version #1.25
APPLICATION NUMBER: US/08/435,149
FILING DATE: 05-MAY-1995

CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
REGISTRATION NUMBER: 36,914
REFERENCE/DOCKET NUMBER: 0989.001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2585
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 323 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-435-149-2

Query Match 29.5%; Score 43; DB 2; Length 323;
Best Local Similarity 42.9%; Pred. No. 72;
Matches 9; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 4 GLYQGVPRAPRGTEARRHYDE 24
DB 3 GLPPDPVNAQPALEGRTSFPE 23

RESULT 19
US-08-313-416A-14
Sequence 14, Application US/08310416A
Patent No. 5679546
GENERAL INFORMATION:
APPLICANT: Jones-Long Ko et al.
TITLE OF INVENTION: CHIMERIC PROTEINS WHICH BLOCK
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2404
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB

US-08-074-108-30
: Sequence 30, Application US/08674168

```

; Patent No. 5804414
; GENERAL INFORMATION:
; APPLICANT: MORIYA, Mika
; APPLICANT: MATSUI, Hiroshi
; APPLICANT: YOKOZAKI, Kenzo
; APPLICANT: HIRANO, Seiko
; APPLICANT: HAYAKAWA, Atsushi
; APPLICANT: SUGIMOTO, Masakazu
; APPLICANT: SUGIMOTO, Masakazu
; TITLE OF INVENTION: METHOD OF AMPLIFYING GENE USING
; TITLE OF INVENTION: ARTIFICIAL TRANSCRIPTION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE # 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/674,168
; FILING DATE: 01-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-166541
; FILING DATE: 30-JUN-1995
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 0-810-0
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-674-168-30

Query Match          29.1%; Score 42.5; DB 1; Length 445;
Best Local Similarity 52.4%; Pred. No. 1.2e+02;
Matches 11; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

OY 2 GEGLYQGV-PRAEPGTEARRH 21
DB 154 GEGKIQDVLRVPGIEATH 174

RESULT 23
US-08-985-908-20
; Sequence 20, Application US/08985908
; Patent No. 6047730.
; GENERAL INFORMATION:
; APPLICANT: MASAYUKI ARAKI, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI NA
; TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; COUNTRY: USA
; ZIP: 22202

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,908
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-325659
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-985-908-20

Query Match          29.1%; Score 42.5; DB 3; Length 445;
Best Local Similarity 52.4%; Pred. No. 1.2e+02;
Matches 11; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

OY 2 GEGLYQGV-PRAEPGTEARRH 21
DB 154 GEGKIQDVLRVPGIEATH 174

RESULT 24
US-08-852-730-5
; Sequence 5, Application US/08852730
; Patent No. 6090597
; GENERAL INFORMATION:
; APPLICANT: HIRANO, MASAKAZU SUGIMOTO, EICHI NAKANO,
; APPLICANT: MASAKO IZUI, ATSUSHI HAYAKAWA, YASUHIKO YOSHIHARA, AND TSUYOSHI
; APPLICANT: NAKAMATSU
; TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER AND NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
; CITY: ARLINGTON
; STATE: VA
; ZIP: 22026
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/852,730
; FILING DATE: 05-07-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-142812
; FILING DATE: 05-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids

```

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-852-730-5

Query Match      29.1%; Score 42.5; DB 3; Length 445;
Best Local Similarity 52.4%; Pred. No. 1.2e+02;
Matches 11; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 2 GEGLYQGV-PRAEPGTEARRH 21
   ||| | | | | | | | | |
Db 154 GEGKIQDLVLRKPGIEATH 174

RESULT 25
US-08-985-916-12
; Sequence 12, Application US/08985916
; Patent No. 6221636
; GENERAL INFORMATION:
; APPLICANT: ATSUSHI HAYAKAWA, MASAKAZU SUGIMOTO, YASUHIKO YOSHIHARA, AND TSUYOSHI
; TITLE OF INVENTION: METHOD FOR PRODUCING L-LYSINE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; COUNTRY: VA
; ZIP: 22152
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 8-325658
; FILING DATE: 05-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-985-916-12

Query Match      29.1%; Score 42.5; DB 4; Length 445;
Best Local Similarity 52.4%; Pred. No. 1.2e+02;
Matches 11; Conservative 1; Mismatches 8; Indels 1; Gaps 1;

QY 2 GEGLYQGV-PRAEPGTEARRH 21
   ||| | | | | | | | | |
Db 154 GEGKIQDLVLRKPGIEATH 174

RESULT 26
US-08-652-265-4
; Sequence 4, Application US/08652265
; Patent No. 6025130
; GENERAL INFORMATION:
; APPLICANT: Thomas, Winston J.
; APPLICANT: Drayna, Dennis T.
; APPLICANT: Feder, John N.
; APPLICANT: Gnirke, Andreas
; APPLICANT: Ruddy, David
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Hereditary Hemochromatosis Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
;
; APPLICANT: Gnirke, Andreas
; APPLICANT: Ruddy, David
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Hereditary Hemochromatosis Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,265
; FILING DATE: 23-MAY-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 17957-000500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-652-265-4

Query Match      28.8%; Score 42; DB 3; Length 348;
Best Local Similarity 46.2%; Pred. No. 1.1e+02;
Matches 12; Conservative 1; Mismatches 5; Indels 8; Gaps 2;

QY 2 GEGLYQG-----VPRAEPGTEARRHY 22
   | | | | | | | | | | | | | |
Db 260 GEGTYQGWITLAMP---PGEEQRTY 282

RESULT 27
US-08-652-265-8
; Sequence 8, Application US/08652265
; Patent No. 6025130
; GENERAL INFORMATION:
; APPLICANT: Thomas, Winston J.
; APPLICANT: Drayna, Dennis T.
; APPLICANT: Feder, John N.
; APPLICANT: Gnirke, Andreas
; APPLICANT: Ruddy, David
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Hereditary Hemochromatosis Gene
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```

CURRENT APPLICATION DATA:
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0056-999
FILING DATE: 23-MAY-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 8907-000500
FILING DATE: 04-APR-1996
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-652-265-8

```

```

Query Match
Best Local Similarity 28.8%; Score 42; DB 3; Length 348;
Matches 12; Conservative 1; Mismatches 5; Indels 8; Gaps 2;

```

```

Oy 2 GEGLYQG-----VPRAEPGTARRHY 22
Db 260 GDGTQGWITLAVP---PGEQRVY 282

```

```

RESULT 28
US-08-834-497A-4

```

```

; Sequence 4, Application US/08834497A
; Patent No. 6140305

```

```

GENERAL INFORMATION:

```

```

APPLICANT: Thomas, Winston J.

```

```

APPLICANT: Drayna, Dennis T.

```

```

APPLICANT: Feder, John N.

```

```

APPLICANT: Gaierke, Andreas

```

```

APPLICANT: Ruddy, David

```

```

APPLICANT: Tsuchihashi, Zenta

```

```

APPLICANT: Wolff, Roger K.

```

```

TITLE OF INVENTION: HEREDITARY HEMOCHROMATOSIS GENE PRODUCTS

```

```

CORRESPONDENCE ADDRESS:

```

```

ADDRESSEE: Pennie & Edmonds LLP

```

```

STREET: 1155 Avenue of the Americas

```

```

CITY: New York

```

```

STATE: New York

```

```

COUNTRY: USA

```

```

ZIP: 10036-2811

```

```

COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: Windows 95

```

```

SOFTWARE: FASTSEQ for Windows Version 2.0b

```

```

CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/834,497A

```

```

FILING DATE: 04-APR-1997

```

```

CLASSIFICATION: 514

```

```

PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US 08/652,265

```

```

FILING DATE: 23-MAY-1996

```

```

CLASSIFICATION: 514

```

```

PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US 08/630,912

```

```

FILING DATE: 04-APR-1996

```

```

CLASSIFICATION: 514

```

```

ATTORNEY/AGENT INFORMATION:

```

```

NAME: Poissant, Brian M.

```

```

REGISTRATION NUMBER: 28,462

```

```

REFERENCE/DOCKET NUMBER: 8907-0056-999

```

```

TELEPHONE: 650-493-4935

```

```

TELEFAX: 650-493-5556

```

```

REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0056-999
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
MOLECULE TYPE: protein
US-08-834-497A-4

```

```

Query Match
Best Local Similarity 46.2%; Score 42; DB 4; Length 348;
Matches 12; Conservative 1; Mismatches 5; Indels 8; Gaps 2;

```

```

Oy 2 GEGLYQG-----VPRAEPGTARRHY 22
Db 260 GDGTQGWITLAVP---PGEQRVY 282

```

```

RESULT 29

```

```

US-08-834-497A-8

```

```

; Sequence 8, Application US/08834497A
; Patent No. 6140305

```

```

GENERAL INFORMATION:

```

```

APPLICANT: Thomas, Winston J.

```

```

APPLICANT: Drayna, Dennis T.

```

```

APPLICANT: Feder, John N.

```

```

APPLICANT: Gaierke, Andreas

```

```

APPLICANT: Ruddy, David

```

```

APPLICANT: Tsuchihashi, Zenta

```

```

APPLICANT: Wolff, Roger K.

```

```

TITLE OF INVENTION: HEREDITARY HEMOCHROMATOSIS GENE PRODUCTS

```

```

CORRESPONDENCE ADDRESS:

```

```

ADDRESSEE: Pennie & Edmonds LLP

```

```

STREET: 1155 Avenue of the Americas

```

```

CITY: New York

```

```

STATE: New York

```

```

COUNTRY: USA

```

```

ZIP: 10036-2811

```

```

COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: Windows 95

```

```

SOFTWARE: FASTSEQ for Windows Version 2.0b

```

```

CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/834,497A

```

```

FILING DATE: 04-APR-1997

```

```

CLASSIFICATION: 514

```

```

PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US 08/652,265

```

```

FILING DATE: 23-MAY-1996

```

```

CLASSIFICATION: 514

```

```

PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US 08/630,912

```

```

FILING DATE: 04-APR-1996

```

```

CLASSIFICATION: 514

```

```

ATTORNEY/AGENT INFORMATION:

```

```

NAME: Poissant, Brian M.

```

```

REGISTRATION NUMBER: 28,462

```

```

REFERENCE/DOCKET NUMBER: 8907-0056-999

```

```

TELEPHONE: 650-493-4935

```

```

TELEFAX: 650-493-5556

```


US-09-503-444A-4

QY 2 GEGLYQG----VPRAEPGTEARRHY 22
|:| || | | | | | | | |

Db 260 GDGTQGWITLVP---PGEORVY 282

RESULT 32

US-08-960-022-8
; Sequence 8, Application US/08960022
; Patent No. 5978837
; GENERAL INFORMATION:
; APPLICANT: MCCOY, John M.
; APPLICANT: Lavalie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 4600 Middleburg Park Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140

COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 01-JUL-1994
; CLASSIFICATION: 514
; PRIORITY INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION SOURCE: Genbank
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-960-022-8

Query Match 28.8%; Score 42; DB 2; Length 350;
Best Local Similarity 47.1%; Pred. NO. 1.1e+02;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 EGLYQGVPRAEPCTEAR 19

Db 41 ENLVRSPSGEPGRKK 57

RESULT 33

US-08-481-435-4
; Sequence 4, Application US/08481435
; Patent No. 6027906
; GENERAL INFORMATION:
; APPLICANT: Balganes, Tanjore S
; APPLICANT: Town, Christine
; TITLE OF INVENTION: No. 6027906el Polypeptides
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York

STATE: New York
COUNTRY: United States
ZIP: 10036-2787
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 10-JUL-1995
; CLASSIFICATION: 435
; PRIORITY INFORMATION:
; REGISTRATION NUMBER: 1103326-151
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; FILING DATE: 24-NOV-1994
; APPLICATION NUMBER: SE 9404072-2
; ATTORNEY/AGENT INFORMATION:
; REGISTRATION NUMBER: 372
; REFERENCE/DOCKET NUMBER: 1103326-151
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 823 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-481-435-4

Query Match 28.8%; Score 42; DB 3; Length 823;
Best Local Similarity 42.9%; Pred. NO. 3e+02;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 2 GEGLYQGVPRAEPCTEARRHY 22

Db 627 GRVLIQSFQERAVPAQAA 647

RESULT 34

US-08-481-435-9
; Sequence 9, Application US/08481435
; Patent No. 6027906
; GENERAL INFORMATION:
; APPLICANT: Balganes, Tanjore S
; APPLICANT: Town, Christine
; TITLE OF INVENTION: No. 6027906el Polypeptides
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 10-JUL-1995
; CLASSIFICATION: 435
; PRIORITY INFORMATION:
; REGISTRATION NUMBER: 372
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; FILING DATE: 24-NOV-1994
; APPLICATION NUMBER: SE 9404072-2

; FILING DATE: 24-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-151
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 836 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; IMMEDIATE SOURCE:
; CLONE: PARC0469 PBP 1B del 8
US-08-481-435-9

Query Match 28.8%; Score 42; DB 3; Length 836;
Best Local Similarity 42.9%; Pred. No. 3e+02;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Oy 2 GEGLYQGVPAEPGTEARRHY 22
|: ||| |:| |
Db 640 GKVLQSFQAEAVPAQAAY 660

RESULT 35
US-08-481-435-7
; Sequence 7, Application US/08481435
; Patent No. 6027906
; GENERAL INFORMATION:
; APPLICANT: Balganes, Tanjore S
; APPLICANT: Town, Christine
; TITLE OF INVENTION: No. 6027906el Polypeptides
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,435
; FILING DATE: 10-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IN 580/MAS/94
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9404072-2
; FILING DATE: 24-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-151
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 844 amino acids
; TYPE: amino acid

; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; IMMEDIATE SOURCE:
; CLONE: PARC0438 PBP 1B QQA
US-08-481-435-7

Query Match 28.8%; Score 42; DB 3; Length 844;
Best Local Similarity 42.9%; Pred. No. 3.1e+02;
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Oy 2 GEGLYQGVPAEPGTEARRHY 22
|: ||| |:| |
Db 648 GKVLQSFQAEAVPAQAAY 668

RESULT 36
US-08-481-435-8
; Sequence 8, Application US/08481435
; Patent No. 6027906
; GENERAL INFORMATION:
; APPLICANT: Balganes, Tanjore S
; APPLICANT: Town, Christine
; TITLE OF INVENTION: No. 6027906el Polypeptides
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,435
; FILING DATE: 10-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IN 580/MAS/94
; FILING DATE: 01-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9404072-2
; FILING DATE: 24-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-151
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 844 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; IMMEDIATE SOURCE:
; CLONE: PARC0468 PBP 1B QQLL
US-08-481-435-8

Query Match 28.8%; Score 42; DB 3; Length 844;
Best Local Similarity 42.9%; Pred. No. 3.1e+02;

```
Matches 9; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
QY 2 GEGYQGVPRAEPTGTEARRHY 22
I: ||| |||
Db 648 GKVLQSFQPAERAVPAQAA 668

RESULT 37
US-08-777-708C-2
; Sequence 2, Application US/08777708C
; Patent No. 6159934
; GENERAL INFORMATION:
; APPLICANT: PESCOVITZ, ORA H.
; TITLE OF INVENTION: USE OF GHRH-RP TO STIMULATE STEM CELL FACTOR PRODUCTION
; REFERENCE: 7037-175/10-62
; CURRENT FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: US 60/008,933
; PRIOR FILING DATE: 1995-12-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: 2 PatentIn Ver. 2.0
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-777-708C-2

Query Match 28.1%; Score 41; DB 4; Length 30;
Best Local Similarity 43.8%; Pred. No. 9.4;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 3 EGYQGVPRAEPTGTEA 18
I: ||| |||
Db 15 ESILQGFPRMKPSADA 30

RESULT 38
US-08-777-708C-23
; Sequence 23, Application US/08777708C
; Patent No. 6159934
; GENERAL INFORMATION:
; APPLICANT: PESCOVITZ, ORA H.
; TITLE OF INVENTION: USE OF GHRH-RP TO STIMULATE STEM CELL FACTOR PRODUCTION
; REFERENCE: 7037-175/10-62
; CURRENT FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: US 60/008,933
; PRIOR FILING DATE: 1995-12-20
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: 2 PatentIn Ver. 2.0
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Mus musculus
US-08-777-708C-23

Query Match 28.1%; Score 41; DB 4; Length 60;
Best Local Similarity 43.8%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 3 EGYQGVPRAEPTGTEA 18
I: ||| |||
Db 45 ESILQGFPRMKPSADA 60

RESULT 39
US-08-562-311-2
; Sequence 2, Application US/08562311
; Patent No. 5401002
; GENERAL INFORMATION:
; APPLICANT: RAZ, AVRAHAM
```

```
; TITLE OF INVENTION: A METHOD OF DETERMINING THE PROBABILITY
; OF METASTASIS IN A CELL SAMPLE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dykema Gossett
; STREET: STE 505 N. Woodward
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: U.S.
; ZIP: 48304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; CURRENT APPLICATION DATA: Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/562,311
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: US/08/188,225
; FILING DATE:
; APPLICATION NUMBER: US 07/681,242
; FILING DATE: 04-APR-1991
; PRIOR APPLICATION DATA: US 07/294,249
; FILING DATE: 01-JUN-1989
; ATOM: KELLY, ROBERT L.
; REGISTRATION NUMBER: 31,843
; REFERENCE/DOCKET NUMBER: 61,686-
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 810-540-0849
; TELEFAX: 810-540-0003
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 250 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-562-311-2

Query Match 28.1%; Score 41; DB 1; Length 250;
Best Local Similarity 53.3%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 4 GYQGVPRAEPTGTEA 18
Db 61 GAYHGAFGAYGAPA 75

RESULT 40
US-08-946-914-10
; Sequence 10, Application US/08946914
; Patent No. 6029915
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; REFERENCE: 7037-175/10-62
; CURRENT FILING DATE: 1996-12-20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-9334
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093
FILING DATE: 09-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 250 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-946-914-10

Query Match 28.1%; Score 41; DB 3; Length 250;

Best Local Similarity 53.3%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 4 GLYQGVPRAPGTEA 18
| | | | | | | |
DB 61 GAYHGAPGAYPGAPA 75

RESULT 41

US-09-286-691-29
Sequence 29, Application US/09286691
Patent No. 6190189

GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
APPLICANT: Chen, Huizhong
TITLE OF INVENTION: Cellulases and Coding Sequences
FILE REFERENCE: 42-96
CURRENT APPLICATION NUMBER: US/09/286,691
CURRENT FILING DATE: 1999-04-05
EARLIER APPLICATION NUMBER: US 60/027,883
EARLIER FILING DATE: 1996-10-04
EARLIER APPLICATION NUMBER: PCT US97/18008
EARLIER FILING DATE: 1997-10-03
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 29
LENGTH: 290
TYPE: PRT
ORGANISM: Streptomyces Ksm-9
US-09-286-691-29

Query Match 28.1%; Score 41; DB 4; Length 290;

Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 11; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 4 GLYQGVPRAPGTEARRHYDEG 25
| | : | | | | | | | |
DB 145 GLAEAVTDANP--EARVYIDVG 164

RESULT 42

US-08-078-683A-3
Sequence 3, Application US/08078683A
Patent No. 5486599
GENERAL INFORMATION:

APPLICANT: Saunders, Scott
APPLICANT: Bernfield, Merton
APPLICANT: Kato, Masato
TITLE OF INVENTION: Construction and Use of Synthetic
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/078,683A
FILING DATE: 17-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: CME-062
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-078-683A-3

Query Match 28.1%; Score 41; DB 1; Length 310;
Best Local Similarity 45.5%; Pred. No. 1.4e+02;
Matches 10; Conservative 2; Mismatches 6; Indels 4; Gaps 1;

QY 2 GEGLYQG----VPRAPGTEAR 19
| | | : | : | | | | |
DB 96 GEGPKEGEAVVLPVEPEGLTAR 117

RESULT 43

US-08-488-199-4
Sequence 4, Application US/08488199
Patent No. 5851993

GENERAL INFORMATION:
APPLICANT: Jalkanen, Markku
APPLICANT: Mali, Markku
TITLE OF INVENTION: Suppression of Tumor Cell Growth By
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
STREET: 1100 New York Ave., NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,199
FILING DATE: 07-JUN-1995

CLASSIFICATION: 514
PRIOR APPLICATION DATA: US 08/258,862
FILING DATE: 13-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Cimbal, Michele A.
REGISTRATION NUMBER: 33,851
TELEPHONE: 202-371-2540
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2540
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-488-199-4

Query Match
Best Local Similarity 28.1%; Score 41; DB 2; Length 310;
Matches 10; Conservative 2; Mismatches 6; Indels 4; Gaps 1;

QY 2 GEGLYQG---VPRAPGTGEAR 19
DB 96 GEGPKGEAVVLPEVEPGLTAR 117

RESULT 44
US-09-032-315-3
Sequence 3, Application US/09032315
Patent No. 5985818
GENERAL INFORMATION:
APPLICANT: Svendsen, Allan
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5985818 of No. 5985818th America, Inc.
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA: Windows Version 2.0
FILING DATE: 27-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Svendsen, Allan
REGISTRATION NUMBER: 36,993
REFERENCE/DOCKET NUMBER: 5200.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 499 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: protein
US-09-032-315-3

Query Match
Best Local Similarity 28.1%; Score 41; DB 2; Length 499;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 6 YQGVPRAPGPT 16
DB 281 YQGVPRAPGPT 291

RESULT 45
US-08-993-318A-3
Sequence 3, Application US/08993318A
Patent No. 5998353
GENERAL INFORMATION:
APPLICANT: Pedersen, Anders
TITLE OF INVENTION: LACCASE MUTANTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5998353 of No. 5998353th America
STREET: 105 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS/MS-DOS
CURRENT APPLICATION DATA: Version #1.0, Version #1.30 (EPO)
FILING DATE: December 18, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pedersen, Anders
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5032.200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 499 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: protein
US-08-993-318A-3

Query Match
Best Local Similarity 28.1%; Score 41; DB 2; Length 499;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 YQGVPRAPGPT 16
DB 281 YQGVPRAPGPT 291

Search completed: June 28, 2001, 11:56:14
Job time: 195 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:56:14 ; Search time 37.02 Seconds
(without alignments)
5.442 Million cell updates/sec

Title: US-09-439-313-573

Perfect score: 45

Sequence: 1 MGSIVQLSQS 10

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	255	4	US-09-071-710-36
2	31	68.9	468	1	US-08-459-287-2
3	31	68.9	468	1	US-08-459-287-3
4	31	68.9	468	4	US-08-764-870-8
5	29	64.4	240	2	US-08-800-198-8
6	29	64.4	240	4	US-09-296-595-8
7	29	64.4	241	1	US-08-235-838-5
8	29	64.4	241	1	US-08-235-838-11
9	29	64.4	241	2	US-08-465-473B-5
10	29	64.4	241	2	US-08-465-473B-11
11	29	64.4	248	2	US-08-887-352B-22
12	29	64.4	248	2	US-08-887-352B-23
13	29	64.4	248	4	US-09-109-207C-22
14	29	64.4	248	4	US-09-109-207C-23
15	29	64.4	249	2	US-08-797-689-18
16	29	64.4	271	2	US-08-400-115-4
17	29	64.4	301	2	US-08-661-052-14
18	29	64.4	301	2	US-08-656-906-25
19	29	64.4	301	4	US-09-217-847-25
20	29	64.4	355	3	US-08-875-811-41
21	29	64.4	355	3	US-08-875-811-49
22	29	64.4	355	3	US-08-875-811-64
23	29	64.4	358	3	US-08-875-811-45
24	29	64.4	358	3	US-08-875-811-51
25	29	64.4	360	3	US-08-875-811-47
26	29	64.4	379	3	US-08-875-811-43
27	29	64.4	553	2	US-08-661-052-16

ALIGNMENTS

RESULT 1

US-09-071-710-36

; Sequence 36, Application US/09071710
; Patent No. 6130043
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GORDON, JULIAN
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; FOR DETECTING DISEASES OF THE PROSTATE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

28	29	64.4	637	1	US-08-235-838-14	Sequence 14, Appl
29	29	64.4	637	1	US-08-235-838-16	Sequence 16, Appl
30	29	64.4	637	2	US-08-465-473B-14	Sequence 14, Appl
31	29	64.4	637	2	US-08-465-473B-16	Sequence 16, Appl
32	29	64.4	711	1	US-08-235-838-7	Sequence 7, Appl
33	29	64.4	711	2	US-08-465-473B-7	Sequence 7, Appl
34	29	64.4	1070	2	US-08-633-770A-2	Sequence 13, Appl
35	28	62.2	241	4	US-08-902-486-13	Sequence 10, Appl
36	28	62.2	321	4	US-08-748-506-10	Sequence 18, Appl
37	28	62.2	321	4	US-08-748-506-18	Sequence 15, Appl
38	28	62.2	496	4	US-08-902-486-15	Sequence 36, Appl
39	28	62.2	612	2	US-08-933-750C-36	Sequence 2, Appl
40	28	62.2	612	4	US-09-234-613-36	Sequence 2, Appl
41	28	62.2	680	1	US-07-674-287B-2	Sequence 4, Appl
42	28	62.2	680	2	US-08-436-900A-2	Sequence 4, Appl
43	28	62.2	682	2	US-08-436-900A-4	Sequence 2, Appl
44	28	62.2	910	4	US-08-460-269C-2	Sequence 2, Appl
45	28	62.2	1487	2	US-08-760-489-2	Sequence 4, Appl
46	28	62.2	1487	2	US-08-760-489-4	Sequence 6, Appl
47	28	62.2	1622	4	US-08-972-927-6	Sequence 22, Appl
48	27	60.0	24	2	US-08-493-235-22	Sequence 21, Appl
49	27	60.0	26	2	US-08-493-235-21	Sequence 6, Appl
50	27	60.0	146	1	US-08-688-609-6	Sequence 3, Appl
51	27	60.0	146	3	US-09-002-832-6	Sequence 3, Appl
52	27	60.0	148	1	US-08-688-609-3	Sequence 3, Appl
53	27	60.0	148	3	US-09-002-832-3	Sequence 1, Appl
54	27	60.0	181	5	PCT-US93-07805-1	Sequence 95, Appl
55	27	60.0	220	4	US-09-004-731-95	Sequence 8, Appl
56	27	60.0	226	4	US-08-749-699-95	Sequence 148, App
57	27	60.0	226	4	US-09-004-731-8	Sequence 148, App
58	27	60.0	226	4	US-08-749-699-8	Sequence 148, App
59	27	60.0	240	1	US-08-488-113B-148	Sequence 148, App
60	27	60.0	240	1	US-08-477-484B-148	Sequence 148, App
61	27	60.0	240	2	US-08-646-360-148	Sequence 148, App
62	27	60.0	240	4	US-08-839-765-148	Sequence 148, App
63	27	60.0	240	4	US-09-136-389-148	Sequence 6, Appl
64	27	60.0	243	1	US-08-133-804-6	Sequence 6, Appl
65	27	60.0	243	1	US-08-461-838-6	

OPERATING SYSTEM: DOS
FASTSEQ FOR Windows Version 2.0
SEQUENCE CHARACTERISTICS:
APPLICATION NUMBER: US/09/071,710
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/850,713
APPLICATION NUMBER:
FILING DATE: 02-MAY-1997
NAME/BACKET: INCHEN
REGISTRATION NUMBER: 35,441
REFERENCE/DOCKET NUMBER: 6083, US.P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847/935-1729
TELEFAX: 847/938-2623
SEQUENCE CHARACTERISTICS:
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 255 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: NO. 6130043e
US-09-071-710-36

Query Match 100.0%; Score 45; DB 4; Length 255;
Best Local Similarity 100.0%; Pred. No. 0.12; 0; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;

QY 1 MGSIVOLS 10
DB 212 MGSIVOLS 221

RESULT 2
US-08-459-287-2 Application US/08459287
Sequence 2, 5686596
GENERAL INFORMATION:
APPLICANT: Mukherjee, Ranjan
TITLE OF INVENTION: HUMAN PEROXISOME PROLIFERATOR ACTIVATED
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
STRANDEDNESS: single
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM compatible
OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
FILING DATE: 08/14/1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
FILING DATE: October 25, 1993
APPLICATION NUMBER: 08/141,500
FILING DATE: October 22, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 204/138
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440

TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-459-287-2

Query Match 68.9%; Score 31; DB 1; Length 468;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLS 10
DB 31 MGSIVOLS 40

RESULT 3
US-08-459-287-3 Application US/08459287
Sequence 3, 5686596
GENERAL INFORMATION:
APPLICANT: Mukherjee, Ranjan
TITLE OF INVENTION: HUMAN PEROXISOME PROLIFERATOR ACTIVATED
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
STRANDEDNESS: single
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM compatible
OPERATING SYSTEM: IBM P.C. DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
FILING DATE: 08/14/1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
FILING DATE: October 25, 1993
APPLICATION NUMBER: 08/141,500
FILING DATE: October 22, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 204/138
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-459-287-3

Query Match 68.9%; Score 31; DB 1; Length 468;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLS 10
DB 31 MGSIVOLS 40

RESULT 4
 US-08-764-870-8
 ; Sequence 8, Application US/08764870
 ; Patent No. 6236946
 ; GENERAL INFORMATION:
 ; APPLICANT: Scanlan, Thomas S
 ; APPLICANT: Baxter, John D
 ; APPLICANT: Fletcher, Robert J
 ; APPLICANT: Wagner, Richard L
 ; APPLICANT: Kushner, Peter J
 ; APPLICANT: Arpeltti, James W
 ; APPLICANT: West, Brian
 ; TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
 ; TITLE OF INVENTION: Binding Domains
 ; NUMBER OF SEQUENCES: 16
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooley Godward
 ; STREET: Five Palo Alto Square, 3000 El Camino Real
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/764,870
 ; FILING DATE: 13-DEC-1996
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/008,540
 ; FILING DATE: 13-DEC-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/008,543
 ; FILING DATE: 13-DEC-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/008,606
 ; FILING DATE: 14-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Nakamura, Jackie N
 ; REGISTRATION NUMBER: 35,966
 ; REFERENCE/DOCKET NUMBER: UCAL-246/01US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (650)843-5000
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 468 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-764-870-8

Query Match 68.9%; Score 31; DB 4; Length 468;
 Best Local Similarity 60.0%; Pred. No. 1.3e+02;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVLSQS 10
 Db 31 MGNIOEISQS 40

RESULT 5
 US-08-800-198-8
 ; Sequence 8, Application US/08800198
 ; Patent No. 5942602
 ; GENERAL INFORMATION:
 ; APPLICANT: WELS, WINFRIED S.
 ; APPLICANT: SCHMIDT, MATHIAS
 ; APPLICANT: VAKALOPOULOU, EVANGELIA

; APPLICANT: SCHNEIDER, DOUGLAS
 ; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MILLEN, WHITE, ZELANO & BRANIGAN, P.C.
 ; STREET: 2200 CLARENDON BLVD. SUITE 1400
 ; CITY: ARLINGTON
 ; STATE: VA
 ; COUNTRY: US
 ; ZIP: 22201
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/800,198
 ; FILING DATE: 13-FEB-1997
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: HAMLET-KING, DIANA
 ; REGISTRATION NUMBER: 33,302
 ; REFERENCE/DOCKET NUMBER: SCH 1576
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-243-6333
 ; TELEFAX: 703-243-6410
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 240 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; FRAGMENT TYPE: internal
 ; US-08-800-198-8

Query Match 64.4%; Score 29; DB 2; Length 240;
 Best Local Similarity 66.7%; Pred. No. 1.6e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVLSQS 10
 Db 133 GSDIQLTQS 141

RESULT 6
 US-09-296-595-8
 ; Sequence 8, Application US/09296595A
 ; Patent No. 6129915
 ; GENERAL INFORMATION:
 ; APPLICANT: WELS, WINFRIED S.
 ; APPLICANT: SCHMIDT, MATHIAS
 ; APPLICANT: VAKALOPOULOU, EVANGELIA
 ; APPLICANT: SCHNEIDER, DOUGLAS
 ; TITLE OF INVENTION: GROWTH FACTOR RECEPTOR ANTIBODIES
 ; FILE REFERENCE: SCH-1576 D1
 ; CURRENT APPLICATION NUMBER: US/09/296,595A
 ; CURRENT FILING DATE: 1999-04-23
 ; EARLIER APPLICATION NUMBER: 08/800,198
 ; EARLIER FILING DATE: 1997-02-13
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 8
 ; LENGTH: 240
 ; TYPE: PRT
 ; ORGANISM: Murine sp.
 ; US-09-296-595-8

Query Match 64.4%; Score 29; DB 4; Length 240;
 Best Local Similarity 66.7%; Pred. No. 1.6e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
Db 133 GSDIQLTQS 141.

RESULT 7
US-08-235-838-5
; Sequence 5, Application US/08235838
; Patent No. 5571894
; GENERAL INFORMATION:
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5571894man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,838
; FILING DATE: TBA
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-235-838-5

Query Match 64.4%; Score 29; DB 1; Length 241;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
Db 134 GSDIQLTQS 142

RESULT 8
US-08-235-838-11
; Sequence 11, Application US/08235838
; Patent No. 5571894
; GENERAL INFORMATION:
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5571894man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 564 Morris Avenue

APPLICANT: Wells, Winfried S.
APPLICANT: Hynes, Nancy E.
APPLICANT: Harwerth, Ina-Maria
APPLICANT: Groner, Bernd
APPLICANT: Hardman, No. 5571894man
APPLICANT: Zwickl, Markus
TITLE OF INVENTION: Recombinant Antibodies Specific for a
TITLE OF INVENTION: Growth Factor Receptor
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/235,838
FILING DATE: TBA
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US 07/828,832
FILING DATE: 31-JAN-1992
APPLICATION NUMBER: GB 91-810079.3
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
MOLECULE TYPE: Protein
US-08-235-838-11

Query Match 64.4%; Score 29; DB 1; Length 241;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
Db 135 GSDIQLTQS 143

RESULT 9
US-08-465-473B-5
; Sequence 5, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 564 Morris Avenue

;
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hesna J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-465-473B-5

Query Match 64.4%; Score 29; DB 2; Length 241;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
|| :||:|
Db 134 GSDIQLTQS 142

RESULT 10
US-08-465-473B-11
; Sequence 11, Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wels, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Groner, Bernd
; APPLICANT: Hardman, No. 5939531man
; APPLICANT: Zwickl, Markus
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 564 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,473B
; FILING DATE: 5 June 1995

;
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/828,832
; FILING DATE: 31-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hesna J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-465-473B-11

Query Match 64.4%; Score 29; DB 2; Length 241;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
|| :||:|
Db 135 GSDIQLTQS 143

RESULT 11
US-08-887-352B-22
; Sequence 22, Application US/08887352B
; Patent No. 5994511
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,352B
; FILING DATE: 03-Jul-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 248 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-887-352B-22

Query Match 64.4%; Score 29; DB 2; Length 248;

Best Local Similarity 66.7%; Pred. No. 1.6e+02; DB 2; Length 248;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
DB 135 GSDIQLTQS 143

RESULT 12 352B-23
US-09-109-207C-22
; Sequence 23, Application US/08887352B
; Patent No. 5994511
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
; TITLE OF INVENTION: Improving Polypeptides
; CURRENT APPLICATION NUMBER: US/09/109,207C
; PRIOR APPLICATION NUMBER: US 60/051,554
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA: US/08/887,352B
FILING DATE: 03-Jul-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Svoboda, Craig G.
REGISTRATION NUMBER: 39,044
REFERENCE/DOCKET NUMBER: P1123
TELEPHONE: 650/225-1489
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 amino acids
TYPE: Amino Acid
ORIGIN: Linear
US-08-887-352B-23

Query Match 64.4%; Score 29; DB 2; Length 248;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
DB 135 GSDIQLTQS 143

RESULT 13
US-09-109-207C-22
; Sequence 23, Application US/09109207C
; Patent No. 6172213
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of Improving Polypeptide
; FILE REFERENCE: P1123R1
; CURRENT APPLICATION NUMBER: US/09/109,207C
; PRIOR APPLICATION NUMBER: US 60/051,554
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

; ORGANISM: Artificial
; FEATURE:
; NAME/KEY: Artificial
; LOCATION: 1-248
; OTHER INFORMATION: sfv sequence derived from MAE11
US-09-109-207C-22

Query Match 64.4%; Score 29; DB 4; Length 248;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
DB 135 GSDIQLTQS 143

RESULT 14
US-09-109-207C-23
; Sequence 23, Application US/09109207C
; Patent No. 6172213
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of Improving Polypeptide
; FILE REFERENCE: P1123R1
; CURRENT APPLICATION NUMBER: US/09/109,207C
; PRIOR APPLICATION NUMBER: US 60/051,554
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

Query Match 64.4%; Score 29; DB 4; Length 248;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
DB 135 GSDIQLTQS 143

RESULT 15
US-08-797-689-18
; Sequence 18, Application US/08797689
; Patent No. 5876969
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; APPLICANT: Gournier, Alain
; APPLICANT: Gournier, Jean-Dominique
; APPLICANT: Jung, Gerard
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: NOVEL BIOLOGICALLY ACTIVE POLYPEPTIDES, COMPOSITION
; TITLE OF INVENTION: PREPARATION THEREOF AND PHARMACEUTICAL COMPOSITION
; NUMBER OF SEQUENCES: 36
; CURRENT APPLICATION NUMBER: US/08/797,689-18
; PRIOR APPLICATION NUMBER: US 60/051,554
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Road, 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19426
; COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: Macintosh
OPERATING SYSTEM: System 7.1
SOFTWARE: Word 5.1 (Patentin)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/797,689
FILING DATE: 31-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,927
FILING DATE: 28-JUL-1994
APPLICATION NUMBER: FR 92/01064
FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00085
FILING DATE: 28-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith Ph.D., Julie K.
REGISTRATION NUMBER: P-38,619
REFERENCE/DOCKET NUMBER: ST92006-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 454-3839
TELEFAX: (610) 454-3808
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 249 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-797-689-18

Query Match 64.4%; Score 29; DB 2; Length 249;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
II :||:II
DB 141 GSNQLTQS 149

RESULT 16
US-08-400-115-4
; Sequence 4, Application US/08400115
; Patent No. 5864019
; GENERAL INFORMATION:
; APPLICANT: KING, David John
; APPLICANT: MOUNTAIN, Andrew
; APPLICANT: OWENS, Raymond John
; APPLICANT: YARRANTON, Geoffrey Thomas
; TITLE OF INVENTION: Multivalent Antigen-Binding Proteins
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,115
; FILING DATE: 06-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/127,136
; FILING DATE: 27-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/842,193
; FILING DATE: 17-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB91/00935

; FILING DATE: 11-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9012995.8
; FILING DATE: 11-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: ISACSON, John P.
; REGISTRATION NUMBER: 33,715
; REFERENCE/DOCKET NUMBER: 040283/0106 CARA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 271 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-400-115-4

Query Match 64.4%; Score 29; DB 2; Length 271;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
II :||:II
DB 144 GSQVQLQQS 152

RESULT 17
US-08-661-052-14
; Sequence 14, Application US/08661052
; Patent No. 5837243
; GENERAL INFORMATION:
; APPLICANT: Yashwant M. Deo
; APPLICANT: Joel Goldstein
; APPLICANT: Robert Graziano
; APPLICANT: Chezian Somasundaram
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED
; TITLE OF INVENTION: OF ANTI-FC RECEPTOR ANTIBODIES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/661,052
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/484,172
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: MXI-043CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-661-052-14

Query Match 64.4%; Score 29; DB 2; Length 301;
Best Local Similarity 66.7%; Pred. NO. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSVIQLSOS 10
DB 153 GSDIQLTQS 161

RESULT 18
US-08-656-906-25
; Sequence 25, Application US/08656906
; Patent No. 620801
; GENERAL INFORMATION:
; APPLICANT: Ferkol Jr., Thomas W.
; APPLICANT: Davis, Pamela B.
; APPLICANT: Assen-Galal
; TITLE OF INVENTION: Serpin Enzyme Complex Receptor -
; NUMBER OF SEQUENCES: 31
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 03-JUN-1996
; PRIORITY NUMBER: 08/656,906
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 03-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 95/25809
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/216,534
; FILING DATE: 23-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-656-906-25
Query Match 64.4%; Score 29; DB 2; Length 301;
Best Local Similarity 66.7%; Pred. NO. 2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GSVIQLSOS 10
DB 125 GSEVQLQOS 133

US-08-875-811-41
; Sequence 41, Application US/08875811
; Patent No. 6045793
; GENERAL INFORMATION:
; APPLICANT: Rybak
; APPLICANT: Newton, Dianne L.
; APPLICANT: Boque, Lluís
; APPLICANT: Wlodawer, Alexander
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:

QY 2 GSVIQLSOS 10
DB 125 GSEVQLQOS 133

RESULT 19
US-09-217-847-25
; Sequence 25, Application US/09217847
; Patent No. 620801
; GENERAL INFORMATION:
; APPLICANT: Ferkol Jr., Thomas W.
; APPLICANT: Davis, Pamela B.
; APPLICANT: Assen-Galal
; TITLE OF INVENTION: Serpin Enzyme Complex Receptor -
; NUMBER OF SEQUENCES: 31
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/217,847
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 08/656,906
; APPLICATION NUMBER: 08/656,906
; FILING DATE: 03-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 95/25809
; FILING DATE: 23-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/216,534
; FILING DATE: 23-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 301 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-09-217-847-25
Query Match 64.4%; Score 29; DB 4; Length 301;
Best Local Similarity 77.8%; Pred. NO. 2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GSVIQLSOS 10
DB 125 GSEVQLQOS 133

RESULT 20
US-08-875-811-41
; Sequence 41, Application US/08875811
; Patent No. 6045793
; GENERAL INFORMATION:
; APPLICANT: Rybak
; APPLICANT: Newton, Dianne L.
; APPLICANT: Boque, Lluís
; APPLICANT: Wlodawer, Alexander
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/875,811
FILING DATE: 19-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/02588
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/011,800
FILING DATE: 21-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Faris, Susan K.
REGISTRATION NUMBER: 41,739
REFERENCE/DOCKET NUMBER: 015280-244100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-875-811-41

Query Match 64.4%; Score 29; DB 3; Length 355;
Best Local Similarity 77.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
|| ||| ||

Db 121 GSEVQLQQS 129

RESULT 21
US-08-875-811-49
; Sequence 49, Application US/08875811
; Patent No. 6045793
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Boque, Lluis
; APPLICANT: Wlodawer, Alexander
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,811
; FILING DATE: 19-FEB-1998
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/02588
FILING DATE: 19-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/011,800
FILING DATE: 21-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Faris, Susan K.
REGISTRATION NUMBER: 41,739
REFERENCE/DOCKET NUMBER: 015280-244100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-875-811-49

Query Match 64.4%; Score 29; DB 3; Length 355;
Best Local Similarity 77.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10

|| ||| ||

Db 121 GSEVQLQQS 129

RESULT 22
US-08-875-811-64
; Sequence 64, Application US/08875811
; Patent No. 6045793
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Boque, Lluis
; APPLICANT: Wlodawer, Alexander
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,811
; FILING DATE: 19-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/02588
; FILING DATE: 19-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,800
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Faris, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-244100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:


```

; LENGTH: 355 amino acids
; MOLECULE TYPE: linear
; TOPOLOGY: linear
; NAME/KEY: Protein
; LOCATION: 1..355
; OTHER INFORMATION: /note= "E6FB[Met-(-1)]SerrOnc"
US-08-875-811-51
Query Match 64.4%; Score 29; DB 3; Length 355;
Best Local Similarity 77.8%; Pred. NO. 2.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSEVOLQS 10
DB 239 GSEVOLQS 247

RESULT 24
US-08-875-811-51
; Sequence No 6045733
; PCT NO 6045733
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Boque, Lluís
; APPLICANT: Wlodawer, Alexander
; APPLICANT: Newton, Dianne L.
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; ZIP: 94111-3834
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30.
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,811
; FILING DATE: 19-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/02588
; FILING DATE: 19-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,800
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Faris, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-2441000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-875-811-51

Query Match 64.4%; Score 29; DB 3; Length 358;
Best Local Similarity 77.8%; Pred. NO. 2.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSEVOLQS 10
DB 239 GSEVOLQS 247

RESULT 25
US-08-875-811-47
; Sequence No 6045733
; PCT NO 6045733
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Boque, Lluís
; APPLICANT: Wlodawer, Alexander
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; ZIP: 94111-3834
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/875,811
; FILING DATE: 19-FEB-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/02588
; FILING DATE: 19-FEB-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/011,800
; FILING DATE: 21-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Faris, Susan K.
; REGISTRATION NUMBER: 41,739
; REFERENCE/DOCKET NUMBER: 015280-2441000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-875-811-45

Query Match 64.4%; Score 29; DB 3; Length 358;
Best Local Similarity 77.8%; Pred. NO. 2.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

;; TITLE OF INVENTION: Recombinant Ribonuclease Proteins
;; NUMBER OF SEQUENCES: 64
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Townsend and Townsend and Crew LLP
;; STREET: Two Embarcadero Center, Eighth Floor
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94111-3834
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/875,811
;; FILING DATE: 19-FEB-1998
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/US97/02588
;; FILING DATE: 19-FEB-1997
;; APPLICATION NUMBER: US 60/011,800
;; FILING DATE: 21-FEB-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Paris, Susan K.
;; REGISTRATION NUMBER: 41,739
;; REFERENCE/DOCKET NUMBER: 015280-244100US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; INFORMATION FOR SEQ ID NO: 47:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 360 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-875-811-47

Query Match 64.4%; Score 29; DB 3; Length 360;
Best Local Similarity 77.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
|| ||| ||
DB 241 GSEVQLQQS 249

RESULT 26
US-08-875-811-43
; Sequence 43, Application US/08875811
; Patent No. 6045793
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Boque, Iliuis
; APPLICANT: Wlodawer, Alexander
; TITLE OF INVENTION: Recombinant Ribonuclease Proteins
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/875,811
;; FILING DATE: 19-FEB-1998
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/US97/02588
;; FILING DATE: 19-FEB-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/011,800
;; FILING DATE: 21-FEB-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Paris, Susan K.
;; REGISTRATION NUMBER: 41,739
;; REFERENCE/DOCKET NUMBER: 015280-244100US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; INFORMATION FOR SEQ ID NO: 43:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 379 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-875-811-43

Query Match 64.4%; Score 29; DB 3; Length 379;
Best Local Similarity 77.8%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
|| ||| ||
DB 263 GSEVQLQQS 271

RESULT 27
US-08-661-052-16
; Sequence 16, Application US/08661052
; Patent No. 5837243
; GENERAL INFORMATION:
; APPLICANT: Yashwant M. Deo
; APPLICANT: Joel Goldstein
; APPLICANT: Robert Graziano
; APPLICANT: Chezian Somasundaram
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/661,052
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/484,172
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: MXI-043CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

LENGTH: 553 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-661-052-16

Query Match
Best Local Similarity 64.4%; Score 29; DB 2; Length 553;
Matches 6; Conservative 2; Mismatches 1; Indels 0;

QY 2 GSIVOLQS 10
DB 153 GSDIQLTQS 161

RESULT 28
US-08-235-838-14
Sequence 14, Application US/08235838
Patent No. 5571894
GENERAL INFORMATION:
APPLICANT: Hynes, Nancy E.
APPLICANT: Harwerth, Ina-Maria
APPLICANT: Harward, No. 5571894man
APPLICANT: Zwickl, Markus
TITLE OF INVENTION: Recombinant Antibodies Specific for a
TITLE OF INVENTION: Growth Factor Receptor
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESS: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/235,838
FILING DATE: TBA
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/828,832
FILING DATE: 31-JAN-1992
PUBLICATION NUMBER: GB 91-810079.3
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 637 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-235-838-14

Query Match
Best Local Similarity 64.4%; Score 29; DB 1; Length 637;
Matches 6; Conservative 2; Mismatches 1; Indels 0;

QY 2 GSIVOLQS 10

DB 165 GSDIQLTQS 173

RESULT 29
US-08-235-838-16
Sequence 16, Application US/08235838
Patent No. 5571894
GENERAL INFORMATION:
APPLICANT: Hynes, Nancy E.
APPLICANT: Harwerth, Ina-Maria
APPLICANT: Groner, Bernd
APPLICANT: Hardman, No. 5571894man
APPLICANT: Zwickl, Markus
TITLE OF INVENTION: Recombinant Antibodies Specific for a
TITLE OF INVENTION: Growth Factor Receptor
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESS: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/235,838
FILING DATE: TBA
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/828,832
FILING DATE: 31-JAN-1992
PUBLICATION NUMBER: GB 91-810079.3
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 637 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-235-838-16

Query Match
Best Local Similarity 66.7%; Score 29; DB 1; Length 637;
Matches 6; Conservative 2; Mismatches 1; Indels 0;

QY 2 GSIVOLQS 10
DB 166 GSDIQLTQS 174

RESULT 30
US-08-465-473B-14
Sequence 14, Application US/08465473B
Patent No. 5939531
GENERAL INFORMATION:
APPLICANT: Wells, Winfried S.
APPLICANT: Hynes, Nancy E.
APPLICANT: Harwerth, Ina-Maria

APPLICANT: Groner, Bernd
APPLICANT: Hardman, No. 5939531man
APPLICANT: Zwickl, Markus
TITLE OF INVENTION: Recombinant Antibodies Specific for a
TITLE OF INVENTION: Growth Factor Receptor
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: NOVARTIS Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA
ZIP: 07901-6940
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,473B
FILING DATE: 5 June 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/828,832
FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 91-810079.3
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Pfeiffer, Hesna J.
REGISTRATION NUMBER: 22,640
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908)522 6940
TELEFAX: (908)522 6955
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 637 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-473B-14

Query Match 64.4%; Score 29; DB 2; Length 637;
Best Local Similarity 66.7%; Pred. No. 4.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 GSIVOLSOS 10
Db 165 GSDIQLTQS 173

RESULT 31
US-08-465-473B-16
Sequence 16, Application US/08/465473B
Patent No. 5939531
GENERAL INFORMATION:
APPLICANT: Hynes, Nancy E.
APPLICANT: Zwickl, Markus
APPLICANT: Hardman, No. 5939531man
APPLICANT: Groner, Bernd
TITLE OF INVENTION: Recombinant Antibodies Specific for a
TITLE OF INVENTION: Growth Factor Receptor
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: NOVARTIS Corporation
STREET: 564 Morris Avenue
CITY: Summit
STATE: New Jersey
COUNTRY: USA

ZIP: 07901-6940
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,473B
FILING DATE: 5 June 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/828,832
FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 91-810079.3
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Pfeiffer, Hesna J.
REGISTRATION NUMBER: 22,640
REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908)522 6940
TELEFAX: (908)522 6955
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 637 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-473B-16

Query Match 64.4%; Score 29; DB 2; Length 637;
Best Local Similarity 66.7%; Pred. No. 4.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 GSIVOLSOS 10
Db 166 GSDIQLTQS 174

RESULT 32
US-08-235-838-7
Sequence 7, Application US/08235838
Patent No. 5571894
GENERAL INFORMATION:
APPLICANT: Hynes, Nancy E.
APPLICANT: Zwickl, Markus
APPLICANT: Hardman, No. 5571894man
APPLICANT: Groner, Bernd
TITLE OF INVENTION: Recombinant Antibodies Specific for a
TITLE OF INVENTION: Growth Factor Receptor
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/235,838
FILING DATE: TBA
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/828,832

```

; FILING DATE: 31-JAN-1992
; APPLICATION NUMBER: GB 91-810079.3
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-233-838-7

```

```

Query Match 64.4%; Score 29; DB 1; Length 711;
Best Local Similarity 66.7%; Pred. NO. 4.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
DB 155 GSDIQLTQS 163

```

```

RESULT 33
US-08-465-473B-7
; Sequence 2; Application US/08465473B
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Wells, Winfried S.
; APPLICANT: Hynes, Nancy E.
; APPLICANT: Harwerth, Ina-Maria
; APPLICANT: Hronek, Norbert
; APPLICANT: Zwickl, Markus
; APPLICANT: No. 5939531mah
; TITLE OF INVENTION: Recombinant Antibodies Specific for a
; TITLE OF INVENTION: Growth Factor Receptor
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NOVARTIS Corporation
; STREET: 246 Morris Avenue
; CITY: Summit
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07901-6940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,473B
; FILING DATE: 5 June 1995
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; INFORMATION FOR SEQ ID NO: 7:
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Pfeiffer, Hessa J.
; REGISTRATION NUMBER: 22,640
; REFERENCE/DOCKET NUMBER: 4-18518/A/CIP/CONT2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)522 6940
; TELEFAX: (908)522 6955
; INFORMATION FOR SEQ ID NO: 7:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 711 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-465-473B-7

```

```

Query Match 64.4%; Score 29; DB 2; Length 711;
Best Local Similarity 66.7%; Pred. NO. 4.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
DB 155 GSDIQLTQS 163

```

```

RESULT 34
US-08-633-770A-2
; Sequence 2; Application US/08633770A
; Patent No. 5939531
; GENERAL INFORMATION:
; APPLICANT: Bojsen, Kirsten
; APPLICANT: Yu, Shukun
; APPLICANT: Kragh, Karsten
; APPLICANT: Christensen, Tove
; APPLICANT: Marcussen, Jan
; APPLICANT: Alnæs, Alf
; TITLE OF INVENTION: ALTERNATIVE PURIFICATION OF 1,4-GLUCAN LYASE FROM A FUNGUS, ITS
; TITLE OF INVENTION: PURIFICATION GENE CLONING AND EXPRESSION IN MICROORGANISMS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,770A
; FILING DATE: July 8, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/03398
; FILING DATE: October 5, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Altman, Daniel E.
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: DY006.001APC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714-760-0404
; TELEFAX: 714-760-9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-633-770A-2

```

```

Query Match 64.4%; Score 29; DB 2; Length 1070;
Best Local Similarity 66.7%; Pred. NO. 7.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
DB 1020 GSIVDVLSQA 1028

```

RESULT 35
US-08-902-486-13
; Sequence 13, Application US/08902486
; Patent No. 6140075
; GENERAL INFORMATION:
; APPLICANT: Russel, David R.
; APPLICANT: Fuller, James T.
; TITLE OF INVENTION: METHOD FOR PRODUCING ANTIBODIES AND
; TITLE OF INVENTION: PROTEIN TOXINS IN PLANT CELLS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,486
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 670513.90261
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 241 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-902-486-13

Query Match 62.2%; Score 28; DB 4; Length 241;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
||:||||

Db 128 GSDIQLQSS 136

RESULT 36
US-08-748-506-10
; Sequence 10, Application US/08748506
; Patent No. 6159707
; GENERAL INFORMATION:
; APPLICANT: Ronnett et al.
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,506
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,751
; FILING DATE: 09-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 74940
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5600
; TELEFAX: 312-616-5700
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 321 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-748-506-10

Query Match 62.2%; Score 28; DB 4; Length 321;
Best Local Similarity 60.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
||:||||

Db 157 MGCIVSLGQT 166

RESULT 37
US-08-748-506-18
; Sequence 18, Application US/08748506
; Patent No. 6159707
; GENERAL INFORMATION:
; APPLICANT: Ronnett et al.
; TITLE OF INVENTION: NOVEL SPERM RECEPTORS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,506
; FILING DATE: 08-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,751
; FILING DATE: 09-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 74940
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-616-5600
; TELEFAX: 312-616-5700
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 321 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-748-506-18

Query Match 62.28; Score 28; DB 4; Length 321;
Best Local Similarity 60.08; Pred No. 3 3e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0;
Gaps 0;

QY 1 MGSIVQLSQS 10
DB 157 MGCIVSLQQT 166

RESULT 38
US-08-902-486-15
; Sequence 15, Application US/08902486
; Patent No. 6140075
; GENERAL INFORMATION:
; APPLICANT: Russell, David R.
; TITLE OF INVENTION: METHOD FOR PRODUCING ANTIBODIES AND
; PROTEIN TOXINS IN PLANT CELLS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; STATE: Madison
; COUNTRY: US
; ZIP: 53701-2113

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA: Version #1.0, Version #1.30
; FILING DATE: US/08/902,486
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 77386
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-902-486-15

Query Match 62.28; Score 28; DB 4; Length 496;
Best Local Similarity 66.7%; Pred No. 5 2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0;
Gaps 0;

QY 2 GSIVQLSQS 10
DB 383 GSDIQLQOS 391

RESULT 39
US-08-933-750C-36
; Sequence 36, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shadman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.

APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; CURRENT APPLICATION DATA: Windows Version 2.0
; FILING DATE: September 23, 1997
; CLASSIFICATION: US/08/933,750C
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 612 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: B21N001
; CLONING: 1832295
US-08-933-750C-36

Query Match 62.28; Score 28; DB 2; Length 612;
Best Local Similarity 77.8%; Pred No. 6 1e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0;
Gaps 0;

QY 2 GSIVQLSQS 10
DB 483 GSAGQLSQS 491

RESULT 40
US-09-234-613-36
; Sequence 36, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shadman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 612 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINON01
; CLONE: 1832295
; US-09-234-613-36

Query Match 62.2%; Score 28; DB 4; Length 612;
Best Local Similarity 77.8%; Pred. No. 6.5e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
|| |||||
DB 483 GSAGLSQS 491

RESULT 41
US-07-674-287B-2
; Sequence 2, Application US/07674287B
; Patent No. 5414076
; GENERAL INFORMATION:
; APPLICANT: Bryan Mark O'Hara
; TITLE OF INVENTION: Gibbon Ape Leukemia
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Karen A. Lowney
; ADDRESS: American Cyanamid Company
; STREET: 1937 West Main Street
; STREET: P.O. Box 60
; CITY: Stamford
; STATE: CT
; COUNTRY: USA
; ZIP: 06904-0060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII converted from IBM DW4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/674,287B
; FILING DATE: 19910325
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lowney, Karen A., Dr.

; REGISTRATION NUMBER: 31,274
; REFERENCE/DOCKET NUMBER: 31,104-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 203 321 2361
; TELEFAX: 203 321 2971
; TELEX: 710 474 4059
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 680 Amino Acid Residues
; TYPE: AMINO ACID SEQUENCE
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: protein
; US-07-674-287B-2

Query Match 62.2%; Score 28; DB 1; Length 680;
Best Local Similarity 55.6%; Pred. No. 7.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
|:|:|:|:
DB 357 GNLVQFSQA 365

RESULT 42
US-08-436-900A-2
; Sequence 2, Application US/08436900A
; Patent No. 5874264
; GENERAL INFORMATION:
; APPLICANT: O'Hara, Bryan M.
; TITLE OF INVENTION: Gibbon Ape Leukemia Virus Receptors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: American Home Products
; STREET: One Campus Drive
; CITY: Parsippany
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07054
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,900A
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Barnhard, Elizabeth M.
; REGISTRATION NUMBER: 31,088
; REFERENCE/DOCKET NUMBER: 31,104-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-683-2158
; TELEFAX: 201-683-4117
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 680 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-436-900A-2

Query Match 62.2%; Score 28; DB 2; Length 680;
Best Local Similarity 55.6%; Pred. No. 7.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
|:|:|:|:
DB 357 GNLVQFSQA 365

RESULT 43
US-08-436-900A-4
Sequence 4, Application US/08436900A
Patent No. 5874264
GENERAL INFORMATION:
APPLICANT: O'Hara, Bryan M.
INVENTOR: O'Hara, Bryan M.
TITLE OF INVENTION: Gibbon Ape Leukemia Virus Receptors
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Home Products
STREET: One Campus Drive
CITY: Parsippany
STATE: New Jersey
COUNTRY: U.S.A.
FILING DATE: 08-MAY-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,104-03
TITLE OF INVENTION:
TELEPHONE: 201-683-2158
TELEFAX: 201-683-4117
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 682 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-900A-4

Query Match 62.2%; Score 28; DB 2; Length 682;
Best Local Similarity 55.6%; Pred. No. 7.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVLSQS 10
DB 360 GNLVQFSQT 368

RESULT 44 269C-2
US-08-460-269C-2
Sequence 2, Application US/08460269C
Patent No. 6197548
GENERAL INFORMATION:
APPLICANT: CLARE, JEFFREY J.
INVENTOR: ROMANOS, MICHAEL A.
TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
STREET: 2200 Clarendon Blvd., Suite 1400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Romanos, Michael A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/008001
TELEPHONE: 619/678-5090
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 148 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

US-08-436-900A-4
Sequence 4, Application US/08436900A
Patent No. 5874264
GENERAL INFORMATION:
APPLICANT: O'Hara, Bryan M.
INVENTOR: O'Hara, Bryan M.
TITLE OF INVENTION: Gibbon Ape Leukemia Virus Receptors
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Home Products
STREET: One Campus Drive
CITY: Parsippany
STATE: New Jersey
COUNTRY: U.S.A.
FILING DATE: 08-MAY-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Barnhard, Elizabeth M.
REGISTRATION NUMBER: 31,088
REFERENCE/DOCKET NUMBER: 31,104-03
TITLE OF INVENTION:
TELEPHONE: 201-683-2158
TELEFAX: 201-683-4117
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 682 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-900A-4

Query Match 62.2%; Score 28; DB 4; Length 910;
Best Local Similarity 56.7%; Pred. No. 9.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVLSQS 10
DB 302 GSVSLAQS 310

RESULT 45
US-08-760-489-2
Sequence 2, Application US/08760489
Patent No. 5830696
GENERAL INFORMATION:
APPLICANT: Short, Jay M.
INVENTOR: Short, Jay M.
TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC ENZYMES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fish & Richardson, P.C.
REGISTRATION NUMBER: 60/008,311
REFERENCE/DOCKET NUMBER: 09010/008001
TELEPHONE: 619/678-5090
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 148 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows Version 2.0
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/760,489
FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,311
FILING DATE: 07-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fish & Richardson, P.C.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/008001
TELEPHONE: 619/678-5090
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 148 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
COMPUTER: IBM PC compatible
OPERATING SYSTEM: internal
SOFTWARE: PatentIn Release #1.0, Version #1.30

Query Match 62.2%; Score 28; DB 2; Length 1487;
Best Local Similarity 44.4%; Pred. No. 1.6e+03;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 MGSIVOLSQ 9
:|:::|
Db 221 LGSVIRLQ 229

Search completed: June 28, 2001, 11:56:15
Job time: 196 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: June 28, 2001, 11:54:17 ; Search time 66.55 Seconds
(without alignments)
9.110 Million cell updates/sec

Title: US-09-439-313-573
Perfect score: 45
Sequence: 1 MGSIVQLSQS 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	100.0	231	21	AA54369
2	45	100.0	255	20	AAW5068
3	45	100.0	255	21	AAW29268
4	45	100.0	553	19	AAW71869
5	45	100.0	553	19	AAW69385
6	45	100.0	553	21	AAW28527
7	45	100.0	553	21	AAW82002
8	33	73.3	3413	19	AAW52849
9	31	68.9	468	16	AAW74053
10	31	68.9	468	20	AAW21633
11	31	68.9	468	20	AAW05470

ALIGNMENTS

RESULT 1
AA54369
ID AA54369 standard; Protein; 231 AA.

XX
AC
AA54369;

DT
06-APR-2000 (first entry)

DE
Amino acid sequence of a prostate cancer-associated protein.

XX
Prostate cancer-associated gene; Incyte clone 1864683; bone cancer;

KW
cell proliferation; cancer; adrenal gland cancer; bladder cancer;

KW
prostate cancer.

12	31	68.9	468	22	AAE20342	Peroxisome prolifer
13	30	66.7	288	21	AAG11352	Arabidopsis thalia
14	30	66.7	288	21	AAG16074	Arabidopsis thalia
15	30	66.7	303	21	AAG11351	Arabidopsis thalia
16	30	66.7	303	21	AAG16073	Arabidopsis thalia
17	30	66.7	425	21	AAG11350	Arabidopsis thalia
18	30	66.7	425	21	AAG16072	Arabidopsis thalia
19	30	66.7	750	21	AAW26599	Synechocystis sp p
20	30	66.7	2559	20	AAW41012	Amino acid sequenc
21	30	66.7	5089	19	AAW52846	A. mediterranei ri
22	29	64.4	36	22	AAW65054	Gene #7 associated
23	29	64.4	155	18	AAW55323	H. pylori ORF hp2e
24	29	64.4	160	20	AAW60166	Human endometrium
25	29	64.4	230	21	AAW42674	Human ORFX ORF2438
26	29	64.4	240	16	AAW85494	Erbb-2-specific sc
27	29	64.4	240	16	AAW85495	SCFV(FWP51). Synt
28	29	64.4	240	17	AAW05133	Single chain antib
29	29	64.4	240	17	AAW05134	Single chain antib
30	29	64.4	240	18	AAW25784	Anti-gp130 antibod
31	29	64.4	240	19	AAW71243	scFv comprising he
32	29	64.4	241	13	AAW26979	Fv heavy chain/lig
33	29	64.4	241	13	AAW26981	FWP51 fusion prote
34	29	64.4	242	22	AAW31426	Protein used for i
35	29	64.4	248	16	AAW77606	Murine 5G1.1M1 scF
36	29	64.4	248	18	AAW01829	scFv(1.1ASML) gene
37	29	64.4	248	20	AAW95668	Mus musculus anti-
38	29	64.4	248	20	AAW95667	Mus musculus anti-
39	29	64.4	248	22	AAW76956	SfV fragment of e2
40	29	64.4	248	22	AAW76957	SfV fragment of e2
41	29	64.4	251	9	AAW80152	Multifunctional pr
42	29	64.4	253	21	AAW28159	Murine anti-EGF-2
43	29	64.4	271	13	AAW20185	Sequence of the sh
44	29	64.4	272	14	AAW34672	26-10 sfv. Synthe
45	29	64.4	272	17	AAW00557	Nematode salivary
46	29	64.4	272	19	AAW43913	Mus musculus antib
47	29	64.4	284	21	AAW38451	Fragment of human
48	29	64.4	285	18	AAW55647	H. pylori ORF 06ep
49	29	64.4	285	18	AAW55573	H. pylori ORF 06ep
50	29	64.4	287	19	AAW98675	H. pylori GHPO 548
51	29	64.4	288	20	AAW82743	Fusion protein PNG
52	29	64.4	289	22	AAW65048	Gene #7 associated
53	29	64.4	301	18	AAW11507	Single chain, huma
54	29	64.4	301	19	AAW37085	Anti-human SC sing
55	29	64.4	301	20	AAW73217	Multispecific sing
56	29	64.4	301	22	AAW61959	Single chain human
57	29	64.4	305	21	AAW59264	Antibody 4H5 H cha
58	29	64.4	305	21	AAW51141	Murine derived pro
59	29	64.4	331	22	AAW94268	H6FtripBscFv(CEA6
60	29	64.4	347	22	AAW64999	Human secreted pro
61	29	64.4	355	18	AAW35129	R. pipiens recombi
62	29	64.4	355	18	AAW35125	R. pipiens recombi
63	29	64.4	358	18	AAW35127	R. pipiens recombi
64	29	64.4	358	18	AAW35130	R. pipiens recombi
65	29	64.4	360	18	AAW35128	R. pipiens recombi

OS Homo sapiens.
 XX Key Location/Qualifiers
 XX Peptide 1..47
 FT Modified-site /note= "signal peptide"
 FT Modified-site 100
 FT Modified-site 142
 FT Modified-site /note= "potential casein kinase II phosphorylation site"
 FT Modified-site 142
 FT Modified-site /note= "potential casein kinase II phosphorylation site"
 FT Domain /note= "potential protein kinase C phosphorylation site"
 FT Domain 188..209
 FT /note= "potential transmembrane domain"
 XX XX
 XX WO9967384-A2.
 XX 29-DEC-1999.
 XX 15-JUN-1999; 99WO-US13524.
 XX 22-JUN-1998; 98US-0102615.
 XX (INCY-) INCYTE PHARM INC.
 XX Walker MG, Volkmath W, Klingler TM, Sprinzak EA;
 XX WPI: 2000-126631/11.
 XX N-PSDB; AAZ45677.
 XX Identifying biomolecules for the diagnosis and treatment of diseases
 XX associated with cell-proliferation -
 XX Claim 10; Page 50-51; 52pp; English.
 XX The present sequence is encoded by a prostate cancer-associated gene,
 XX whose cDNA is represented by Incyte clone number 1864683. The cDNA
 XX sequence is used in the method of the invention. The specification
 XX describes a method for identifying biomolecules for the diagnosis and
 XX treatment of prostate cancer, comprising examining polynucleotides,
 XX comprising examining polynucleotides, consisting of prostate cancer-
 XX specific genes, and genes of unknown function, expressed in cDNA
 XX libraries. The patterns of both gene sets are compared to identify
 XX genes of unknown function with similar expression patterns to the
 XX prostate cancer-specific genes. The biomolecules identified by the
 XX method form pharmaceutical compositions useful for the diagnosis and
 XX treatment of prostate cancer. The biomolecules include, but are not
 XX limited to, the prostate gland, bladder and bone, but especially
 XX include cancer of the prostate gland, bladder and bone, but especially
 XX prostate cancer. The method may also be applied using other disease-
 XX specific genes. The prostate cancer-specific genes facilitate the
 XX diagnosis and treatment of cell proliferation disorders.

KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
 KW metastasis; prostatic intraepithelial neoplasia; PIN; cancer;
 KW drug screening; gene therapy.
 XX Homo sapiens.
 XX WO9850567-A1.
 XX 12-NOV-1998.
 XX 01-MAY-1998; 98WO-US08930.
 XX 02-MAY-1997; 97US-0850713.
 XX (ABBO) ABBOTT LAB.
 XX Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
 XX Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;
 XX Roberts-Rapp L, Russell JC, Stroupe SD;
 XX WPI: 1999-034731/03.
 XX N-PSDB; AAV71181.
 XX New isolated prostate-specific polynucleotides - used to develop
 XX products for the diagnosis and treatment of prostate diseases, e.g.
 XX benign hyperplasia, prostatic or prostate cancer
 XX Claim 17; Pages 99-100; 122pp; English.
 XX The present sequence is encoded by the consensus sequence for a PS108
 XX gene. The sequence is a 255 AA polynucleotide. The sequence is
 XX the sequence of the PS108 gene-specific. They are used in the method
 XX of the invention. The specification describes a method for detecting the
 XX presence of a target PS108 polynucleotide in a test sample. The method
 XX comprises contacting the test sample with at least 1 PS108-specific
 XX polynucleotide or complement, and detecting the presence of the target
 XX polynucleotide. The products can be used for detecting, diagnosing,
 XX predicting, or treating one or more of the diseases or conditions of
 XX the prostate such as benign prostatic hyperplasia (BPH),
 XX prostatic intraepithelial neoplasia (PIN) and cancer. In
 XX particular the products can be used in drug screening and gene therapy.

Query Match 100.0%; Score 45; DB 20; Length 255;
 Best Local Similarity 100.0%; Pred. NO. 0.084;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MGSIVQLSQS 10
 DB 212 mgsivqlsqds 221
 RESULT 3
 ID AAB29268
 XX AAB29268 standard; Protein: 255 AA.
 XX AAB29268;
 XX 07-FEB-2001 (first entry)
 XX Human prostate-related PS108 partial protein sequence SEQ ID NO: 36.
 XX Human; prostate cancer; PS108; antibody; tumour; metastasis.
 XX Homo sapiens.
 XX OS
 XX US6130043-A.
 XX 10-OCT-2000.

```
PF 01-MAY-1998; 98US-0071710.
XX
PR 02-MAY-1997; 97US-0850713.
XX
XX (ABBO ) ABBOTT LAB.
XX
PI Cohen M, Colpitts TL, Friedman PM, Gordon J, Granados EN;
PI Billing-Medel PA, Klass MR, Roberts-Rapp L, Stroupe SD, Yu H;
PI Kratochvil JD, Russell JC, Hodges SC;
XX
DR WPI: 2000-655655/53.
XX
PT Methods for detecting target prostate-specific polynucleotides or
PT diseases of the prostate (e.g. prostate cancer), comprising detecting
PT the presence of any of PS108 nucleic acid sequences in a test sample -
XX
XX Example 1; Column 85-88; 55pp; English.
XX
CC The present invention is related to a number of partial coding and
CC protein sequences for the human prostate tissue protein PS108. These
CC sequences can be used in the diagnosis and prognosis of prostate
CC diseases, particularly prostate cancer. They can also be used to produce
CC antibodies which can be used in treatment. The present sequence is one
CC of the PS108 partial protein sequences.
XX
SQ Sequence 255 AA;

Query Match 100.0%; Score 45; DB 21; Length 255;
Best Local Similarity 100.0%; Pred. No. 0.084;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 212 mgsivqlsqqs 221

RESULT 4
AAW71869
ID AAW71869 standard; Protein; 553 AA.
XX
AC AAW71869;
XX
DT 06-JAN-1999 (first entry)
XX
DE Amino acid encoded by prostate tumour clone Ll-12.
XX
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
XX
OS Homo sapiens.
XX
PN W09837093-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US03492.
XX
PR 09-FEB-1998; 98US-0020956.
PR 25-FEB-1997; 97US-0806099.
PR 01-AUG-1997; 97US-0904804.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Xu J;
XX
DR WPI: 1998-609886/51.
DR N-PSDB; AAV61201.
XX
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
XX
XX Example 1; Page 82-84; 130pp; English.

CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a
CC vaccine for the treatment of prostate cancer. The immunogen was
CC isolated from a prostate tumour cDNA library obtained by subtracting
CC a prostate tumour cDNA expression library with a normal tissue cDNA
CC library.
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 45; DB 19; Length 553;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 510 mgsivqlsqqs 519

RESULT 5
AAW69385
ID AAW69385 standard; Protein; 553 AA.
XX
AC AAW69385;
XX
DT 08-DEC-1998 (first entry)
XX
DE Prostate tumour specific gene clone Ll-12 protein.
XX
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy.
XX
OS Homo sapiens.
XX
PN W09837418-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US03690.
XX
PR 09-FEB-1998; 98US-0904809.
PR 25-FEB-1997; 97US-0806596.
PR 01-AUG-1997; 97US-0904809.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Xu J;
XX
DR WPI: 1998-480805/41.
DR N-PSDB; AAV58586.
XX
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
XX
PS Example 1; Page 87-89; 141pp; English.
XX
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 45; DB 19; Length 553;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 MGSIVQLSOS 10
Db 510 mgsivqlsqs 519

RESULT 6

AA28527
ID AAB28527 standard; Protein: 553 AA.
AC AAB28527;
XX DT 07-FEB-2001 (first entry)
XX Protein encoded by human breast tumour cDNA clone p5019.
DE DE
XX Human; breast tumour antigen; cytostatic; immunotherapy;
KW Breast cancer; vaccine.
XX Homo sapiens.
OS
XX WO200061756-A2.
PN
XX 19-OCT-2000.
XX
XX 10-APR-2000; 2000WO-US09688.
PF
XX 09-APR-1999; 99US-0288950.
PR
XX 02-JUL-1999; 99US-0346327.
XX
XX (CORI-) CORIXA CORP.
PA
XX Reed SG, Xu J, Dillon DC;
XX WPI: 2000-638568/61.
DR N-PSDB; AAC79473.
XX
XX A novel isolated polypeptide comprising an immunogenic portion of a
PT breast cancer protein useful in the detection and treatment of breast
PT cancer.
XX
XX Claim 2; Page 92-93; 95pp; English.
XX
XX The present sequence is encoded by a cDNA sequence which was isolated
CC from a breast tumour cDNA library. It is provided in a specification
CC relating to compounds for immunotherapy and diagnosis of breast cancer.
CC Breast tumour antigens and vaccines are described.
CC The present sequence is a portion of a pharmaceutical composition to be used in the
CC treatment of breast cancer. Proliferated T cells and incubated antigen
CC presenting cells are also required. The polypeptides and polynucleotides
CC may also be used to produce a vaccine.
XX
XX Sequence 553 AA:

Query Match 100.0%; Score 45; DB 21; Length 553;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGSIVQLSOS 10
Db 510 mgsivqlsqs 519

RESULT 7

AA28202
ID AAY82002 standard; Protein: 553 AA.
XX
XX AAY82002;
AC
XX 13-JUN-2000 (first entry)
XX Human immunogenic prostate tumour protein sequence SEQ ID NO:113.
DE
XX

KW

Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
immunogenic; cytostatic; vaccine.
Homo sapiens.
WO200004149-A2.
27-JAN-2000.
14-JUL-1999; 99WO-US15838.
14-JUL-1998; 98US-0115453.
14-JUL-1998; 98US-0116134.
23-SEP-1998; 98US-0159822.
15-JAN-1999; 99US-0232149.
15-JAN-1999; 99US-0232880.
09-APR-1999; 99US-0288946.
(CORI-) CORIXA CORP.
Dillon DC, Harlocker SL, Yuqiu J, Xu J, Mitcham JL;
WPI: 2000-171268/15.
New polypeptide useful for treating and diagnosing prostate cancer
comprises an immunogenic portion of prostate tumor protein -
Claim 3; Page 138-139; 263pp; English.
The present invention describes isolated polypeptides, comprising an
immunogenic portion of a prostate tumour protein (PTP). The polypeptides
and polynucleotides encoding them have cytostatic activity and can be
used in vaccines and in gene therapy. The polypeptides and
polynucleotides encoding them, antigen presenting cells which express
the polypeptides, antibodies against polypeptides, polynucleotides and
comprising an immunogenic portion of prostate tumor protein, can be used
or anti-idiotypic antibodies for passive immuno therapy. A portion of
the polynucleotides encoding the polypeptides can be used as a probe or
to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
AA82000 to AAY82020 represent sequences used in the exemplification of
the present invention.
Sequence 553 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PD 26-FEB-1998.
 XX
 XX PF 18-AUG-1997; 97WO-EP04495.
 XX
 XX PR 20-AUG-1996; 96EP-0810551.
 XX
 XX PA (NOVS) NOVARTIS AG.
 XX
 XX PI Engel N, Schupp T, Toupet C;
 XX
 XX DR WPI; 1998-169172/15.
 XX
 XX DR N-PSDB; AAV21187.
 XX
 XX AMycolatopsis mediterranei rifamycin synthesis gene cluster - used
 PT to produce rifamycin and rifamycin analogues
 XX
 XX PS Claim 6; Page 170-187; 205pp; English.
 XX
 XX CC The present sequence represents a Amycolatopsis mediterranei rifamycin
 CC synthesis gene cluster ORF E protein from the present invention. The
 CC DNA fragment comprises a DNA region involved directly or indirectly
 CC in the gene cluster responsible for rifamycin synthesis, including
 CC their function in connection with rifamycin biosynthesis, qualify
 CC as constituents of this rifamycin gene cluster, and functional
 CC fragments, derivatives or constituents of these. The Amycolatopsis
 CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
 CC for producing rifamycin, rifamycin analogues or precursors. It can also
 CC be used for inactivating or modifying genes involved in ansamycin or
 CC rifamycin biosynthesis. The DNA can be used for constructing mutant
 CC actinomycetes strains from which the natural rifamycin or ansamycin
 CC biosynthesis gene cluster has been partly or completely deleted. The
 CC DNA fragment can be used for assembling a library of polypeptide
 CC synthetases, which can be used for assembling a library of polypeptides.
 CC A hybridisation probe of the invention can be used for identifying DNA
 CC fragments involved in the biosynthesis of ansamycins.
 XX
 XX Sequence 3413 AA;
 SQ

Query Match 73.3%; Score 33; DB 19; Length 3413;
 Best Local Similarity 60.0%; Pred. No. 5.4e+02;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 MGSIVQLSQS 10
 : : : : :
 Db 2588 lgavvqlpqs 2597

RESULT 9
 AAR74053
 ID AAR74053 standard; Protein; 468 AA.
 XX
 XX AC AAR74053;
 XX
 XX DT 24-OCT-1995 (first entry)
 XX
 XX DE Human peroxisome proliferator activated receptor.
 XX
 XX KW PPAR; hyperlipidaemia; atherosclerosis.
 XX
 XX OS Homo sapiens.
 XX
 XX PN W09511974-A.
 XX
 XX PD 04-MAY-1995.
 XX
 XX PF 19-OCT-1994; 94WO-US11897.
 XX
 XX PR 26-OCT-1993; 93US-0143215.
 XX
 XX PR 22-OCT-1993; 93US-0141500.
 XX
 XX PA (LIGA-) LIGAND PHARM INC.

XX Mukherjee R;
 PI
 XX WPI; 1995-178867/23.
 DR N-PSDB; AAQ92081.
 XX
 XX PT New human peroxisome proliferator activated receptor DNA - used
 PT to screen for agents for the treatment of hyperlipidaemia and
 PT diseases such as atherosclerosis
 XX
 XX PS Claim 3; Fig 1; 19pp; English.
 XX
 XX CC The sequence is that of the human peroxisome proliferator
 CC activated receptor (PPAR) isolated from a human liver cDNA library
 CC using a probe based on rat PPAR DNA. The human PPAR can be used
 CC for screening cpds. for improved pharmacological profiles for the
 CC treatment of hyperlipidaemia with higher potency, efficacy and fewer
 CC side effects. It can be used to screen cpds. active as primary
 CC endogenous inducers of the human PPAR. Also PPAR can be used for
 CC establishing the tissue specific expression pattern of human PPAR to
 CC aid treatment of diseases such as atherosclerosis.
 XX
 XX SQ Sequence 468 AA;
 SQ

Query Match 68.9%; Score 31; DB 16; Length 468;
 Best Local Similarity 60.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 MGSIVQLSQS 10
 : : : : :
 Db 31 mgniqueisqs 40

RESULT 10
 AAY21633
 ID AAY21633 standard; protein; 468 AA.
 XX
 XX AC AAY21633;
 XX
 XX DT 11-AUG-1999 (first entry)
 XX
 XX DE Ligand binding domain of nuclear receptor hPPARalpha.
 XX
 XX KW Thyroid hormone receptor; aromatic compound; ligand binding domain;
 KW alpha-glucophosphate dehydrogenase; cardiac; obesity; triglyceride;
 KW plasma cholesterol; anti-hypertriglyceridemic; atherosclerosis; GPDH;
 KW thyroid hormone replacement therapy; nuclear receptor.
 XX
 XX OS Homo sapiens.
 XX
 XX PN W09926965-A2.
 XX
 XX PD 03-JUN-1999.
 XX
 XX PF 25-NOV-1998; 98WO-US25296.
 XX
 XX PR 26-NOV-1997; 97US-0980115.
 XX
 XX PA (REGC) UNIV CALIFORNIA.
 XX
 XX PI Aprilletti JW, Baxter JD, Fletterick RJ, Kushner PJ;
 PI Scanlan TS, Shiau AK, Wagner RL, West BL;
 XX
 XX DR WPI; 1999-357810/30.
 XX
 XX PT Modulating activity of a thyroid hormone receptor
 XX
 XX PS Disclosure; Fig 3I-R; 447pp; English.
 XX
 XX CC The invention relates to a method for modulating activity of a thyroid
 CC hormone receptor that comprises administration of an aromatic compound
 CC which fits spatially and preferentially into a thyroid hormone ligand

CC binding domain. The aromatic compound (of a specified formula) can be
CC the alpha-glucosylphosphate dehydrogenase (GPDH) levels, at
CC levels which do not significantly modify cardiac GPDH levels and are
CC indicated in the treatment of obesity. The compound also lower total
CC plasma cholesterol and triglyceride levels and can be used as anti-
CC hypertriglyceridemic agents. The compound may also be used for treating
CC atherosclerosis and may be indicated in thyroid hormone replacement
CC therapy in patients with compromised cardiac function. Several
CC members of the nuclear receptor superfamily.
CC
XX Sequence 468 AA;

Query Match 68.9%; Score 31; DB 20; Length 468;
Best Local Similarity 60.0%; Pred No. 1.5e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 1 MGSIVOLSQS 10
DB 31 mgnlqisqsgs 40
||| :|||

RESULT 11
AAV05470 ID
ID AAY05470 standard; Protein: 468 AA.
XX
AC AAY05470;
XX
XX 07-JUL-1999 (first entry)
XX Human PPAR-alpha protein sequence.
XX
XX Nuclear receptor agonist; antagonist; identification; PPAR;
XX peroxisome proliferator activated receptor.
XX
XX Homo sapiens.
XX
XX WO9918124-A1.
XX
XX 15-APR-1999.
XX
XX 06-OCT-1998; 98WO-US21049.
XX
XX 07-OCT-1997; 97US-0061385.
XX
XX (MERI) MERCK & CO INC.
XX
XX Cummings RT, Hermes JD, Moller DE, Zhou G;
XX
XX WPI: 1999-263998/22.
XX
XX N-PSDB; AAX36521.
XX
XX Identifying nuclear receptor agonists and antagonists
XX
XX Disclosure; Fig 8a; 60pp; English.

XX This sequence is the human peroxisome proliferator activated
XX receptor-alpha (PPAR-alpha). method for identifying nuclear receptor
XX The invention relates to a method for identifying nuclear receptor
XX agonists and antagonists. The method involves measuring fluorescent resonance energy
XX transfer between fluorescently-labelled nuclear receptors and
XX co-activators. The method can be used for identifying agonists and
XX antagonist of nuclear receptors.
XX
XX Sequence 468 AA;

Query Match 68.9%; Score 31; DB 20; Length 468;
Best Local Similarity 60.0%; Pred No. 1.5e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 1 MGSIVOLSQS 10

DB 31 mgnlqisqsgs 40
||| :|||
RESULT 12
AAB20342 ID
ID AAB20342 standard; Protein: 468 AA.
XX
AC AAB20342;
XX
XX 29-MAY-2001 (first entry)
XX
XX Peroxisome proliferator-activated receptor alpha.
XX
XX Peroxisome proliferator-activated receptor alpha; PPAR-alpha;
XX human; steroid hormone receptor; lipid metabolism;
XX low density lipoprotein; cholesterol; apolipoprotein;
XX atherosclerosis; hyperapobetalipoproteinemia;
XX coronary heart disease; diagnosis; prognosis; therapy.
XX
XX Homo sapiens.
XX
XX WO200120037-A2.
XX
XX 22-MAR-2001.
XX
XX 14-SEP-2000; 2000WO-US25124.
XX
XX 17-SEP-1999; 99US-0154736.
XX
XX (COMP-) COMPLEXE HOSPITALIER SAGAMIE.
XX (UYMC-) UNIV MCGILL.
XX
XX Hudson TJ, Gaudet D, Vohl M, Brewer C, Morgan K;
XX
XX WPI: 2001-244828/25.
XX
XX N-PSDB; AAF30517.
XX
XX
XX Claim 21; Fig 3; 52pp; English.

XX The present sequence is that of human peroxisome
XX proliferator-activated receptor alpha (PPAR-alpha), a member of
XX the steroid hormone receptor superfamily that is involved in the
XX control of cellular lipid utilization. A L162V mutation has been
XX identified in the protein which is associated with a mutation
XX with the conserved DNA-binding or C domain of the protein. The
XX presence of the rare L162 allele is associated with a greater
XX likelihood of an individual having, or of developing, a disorder
XX of lipid metabolism, especially high low density lipoprotein
XX cholesterol levels, atherosclerosis, hyperapobetalipoproteinemia
XX and/or coronary heart disease. PPAR-alpha polypeptides and the
XX polynucleotides are used in the method of identifying the
XX metabolism and for treating such a disorder (by administering a
XX functional portion of PPAR-alpha or a nucleic acid encoding it).
XX
XX Sequence 468 AA;

Query Match 68.9%; Score 31; DB 22; Length 468;
Best Local Similarity 60.0%; Pred No. 1.5e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 1 MGSIVOLSQS 10
DB 31 mgnlqisqsgs 40
||| :|||

RESULT 13
AAG11352
ID AAG11352 standard; Protein; 288 AA.
XX AC AAG11352;
XX DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 10036.
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX PN EPI033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 18-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

```
PR 20-AUG-1999; 99US-0149929;
PR 23-AUG-1999; 99US-0149930;
PR 26-AUG-1999; 99US-0150566;
PR 27-AUG-1999; 99US-0150884;
PR 27-AUG-1999; 99US-0151065;
PR 27-AUG-1999; 99US-0151066;
PR 27-AUG-1999; 99US-0151067;
PR 30-AUG-1999; 99US-0151360;
PR 31-AUG-1999; 99US-0151361;
PR 01-SEP-1999; 99US-0151438;
PR 01-SEP-1999; 99US-0151930;
PR 07-SEP-1999; 99US-0152363;
PR 10-SEP-1999; 99US-0153070;
PR 13-SEP-1999; 99US-0153758;
PR 16-SEP-1999; 99US-0154038;
PR 20-SEP-1999; 99US-0154779;
PR 22-SEP-1999; 99US-0155139;
PR 23-SEP-1999; 99US-0155486;
PR 24-SEP-1999; 99US-0155659;
PR 28-SEP-1999; 99US-0156458;
PR 28-SEP-1999; 99US-0156459;
PR 04-OCT-1999; 99US-0157119;
PR 05-OCT-1999; 99US-0157173;
PR 06-OCT-1999; 99US-0157865;
PR 07-OCT-1999; 99US-0158029;
PR 08-OCT-1999; 99US-0158232;
PR 12-OCT-1999; 99US-0158369;
PR 13-OCT-1999; 99US-0158371;
PR 13-OCT-1999; 99US-0159291;
PR 13-OCT-1999; 99US-0159295;
PR 14-OCT-1999; 99US-0159329;
PR 14-OCT-1999; 99US-0159330;
PR 14-OCT-1999; 99US-0159331;
PR 14-OCT-1999; 99US-0159637;
PR 18-OCT-1999; 99US-0159786;
PR 21-OCT-1999; 99US-0160741;
PR 21-OCT-1999; 99US-0160767;
PR 21-OCT-1999; 99US-0160768;
PR 21-OCT-1999; 99US-0160770;
PR 21-OCT-1999; 99US-0160814;
PR 22-OCT-1999; 99US-0160815;
PR 22-OCT-1999; 99US-0160881;
PR 22-OCT-1999; 99US-0160981;
PR 22-OCT-1999; 99US-0160989;
PR 25-OCT-1999; 99US-0161404;
PR 25-OCT-1999; 99US-0161405;
PR 25-OCT-1999; 99US-0161406;
PR 26-OCT-1999; 99US-0161329;
PR 26-OCT-1999; 99US-0161361;
PR 26-OCT-1999; 99US-0161361;
PR 28-OCT-1999; 99US-0161920;
PR 28-OCT-1999; 99US-0161992;
PR 28-OCT-1999; 99US-0161993;
PR 29-OCT-1999; 99US-0162142;

Query Match:
Best Local Similarity 66.7%; Score 30; DB 21; Length 288;
Matches 7; Conservative 77.8%; Pred No. 1.4e+02;
Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
DB 142 GSIVALSQS 150

RESULT 14
ID AAG16074 standard; Protein: 288 AA.
XX AAG16074;
XX AAG16074;
DI 17-OCT-2000 (first entry)
```

XX Arabidopsis thaliana protein fragment SEQ ID NO: 16574.
DX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX EP1033405-A2.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 25-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0126942.
PR 02-APR-1999; 99US-0128714.
PR 08-APR-1999; 99US-0129845.
PR 16-APR-1999; 99US-0130077.
PR 19-APR-1999; 99US-0130449.
PR 21-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130849.
PR 28-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 30-APR-1999; 99US-0132484.
PR 04-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 06-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137328.
PR 03-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.

```
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 66.7%; Score 30; DB 21; Length 288;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVOLSQS 10
Db 142 ggivalsqs 150

RESULT 15
AAG11351
ID AAG11351 standard; Protein; 303 AA.
XX
AC AAG11351;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 10035.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
```



```
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 66.7%; Score 30; DB 21; Length 303;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GSIVOLSQS 10
   | | | | |
Db 157 ggvaialsqs 165

RESULT 16
AAG16073
ID AAG16073 standard; Protein; 303 AA.
XX
AC AAG16073;
XX
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 16573.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.

XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
```

Query Match	Best Local Similarity	Score	DB	Length
Matches	7; Conservative	56.7%	30; 21; DB	303;
29-SEP-1999;	99US-0156596;			
04-OCT-1999;	99US-0157117;			
05-OCT-1999;	99US-0157753;			
06-OCT-1999;	99US-0157865;			
07-OCT-1999;	99US-0158029;			
08-OCT-1999;	99US-0158232;			
12-OCT-1999;	99US-0158369;			
13-OCT-1999;	99US-0159234;			
13-OCT-1999;	99US-0159295;			
14-OCT-1999;	99US-0159329;			
14-OCT-1999;	99US-0159330;			
14-OCT-1999;	99US-0159331;			
14-OCT-1999;	99US-0159637;			
14-OCT-1999;	99US-0159638;			
14-OCT-1999;	99US-0160741;			
21-OCT-1999;	99US-0160767;			
21-OCT-1999;	99US-0160768;			
21-OCT-1999;	99US-0160770;			
21-OCT-1999;	99US-0160814;			
21-OCT-1999;	99US-0160845;			
22-OCT-1999;	99US-0160981;			
22-OCT-1999;	99US-0160989;			
25-OCT-1999;	99US-0161404;			
25-OCT-1999;	99US-0161405;			
25-OCT-1999;	99US-0161406;			
26-OCT-1999;	99US-0161359;			
26-OCT-1999;	99US-0161360;			
26-OCT-1999;	99US-0161361;			
28-OCT-1999;	99US-0161920;			
28-OCT-1999;	99US-0161992;			
28-OCT-1999;	99US-0161993;			
28-OCT-1999;	99US-0161993;			
28-OCT-1999;	99US-0162142;			
Query Match	Best Local Similarity	Score	DB	Length
Matches	7; Conservative	77.8%	30; 21; DB	303;
2 GSIVOLSQS 10				
157 ggvalseqs 165				
RESULT 17				
AGL1350 17				
ID AGL1350 standard; Protein; 425 AA.				
AC AAG11350;				
XX XX				
DT 17-OCT-2000 (first entry)				
XX XX				
XX Arabidopsis thaliana protein fragment SEQ ID NO: 10034.				
XX Protein identification; signal transduction pathway; metabolic pathway;				
KW hybridisation assay; genetic mapping; gene expression control; promoter;				
KW termination sequence.				
XX Arabidopsis thaliana.				
OS Arabidopsis thaliana.				
XX EPI033405-A2.				
XX EPI033405-A2.				
PD 06-SEP-2000.				
XX 25-FEB-2000; 2000EP-0301439.				
XX 25-FEB-1999;				
XX 99US-0121825.				
XX 09-MAR-1999;				
XX 99US-0123548.				
XX 23-MAR-1999;				
XX 99US-0125788.				
XX 29-MAR-1999;				
XX 99US-0126264.				
XX 99US-0126785.				

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

```
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159339.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 16-OCT-1999; 99US-0159538.
PR 16-OCT-1999; 99US-0159539.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160988.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161407.
PR 26-OCT-1999; 99US-0161329.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.
PR 29-OCT-1999; 99US-0162144.

Query Match 66.7%; Score 30; DB 21; Length 425;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 GSIVOLSOS 10
Db 279 gsvivalsgs 287

RESULT 18
AAG16072
ID AAG16072 standard; Protein; 425 AA.
XX
XX AAG16072;
XX
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 16572.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridization assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 29-MAR-1999; 99US-0126785.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130079.
XX 23-APR-1999; 99US-0130113.
XX 23-APR-1999; 99US-0130310.

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132487.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134271.
PR 14-MAY-1999; 99US-0134271.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138447.
PR 10-JUN-1999; 99US-0138447.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140324.
PR 23-JUN-1999; 99US-0140324.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0142155.
PR 02-JUL-1999; 99US-0142155.
PR 06-JUL-1999; 99US-0142300.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144082.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
```


PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 28-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 66.7%; Score 30; DB 21; Length 425;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
| || |||||
Db 279 ggivalsqs 287

RESULT 19

AAB26599
ID AAB26599 standard; protein; 750 AA.

XX
XX AAB26599;
DT 01-FEB-2001 (first entry)
XX
DE Synecocystis sp phytochrome-related gene Cph8.
KW Phytochrome; fluorescent apophytochrome; bilin; Cph.
XX
OS Synecocystis sp.

PN WQ200056355-A1.

PD 28-SEP-2000.

PF 14-MAR-2000; 2000WO-US06607.

PR 19-MAR-1999; 99US-0272809.

PA (REGC) UNIV CALIFORNIA.

XX Lagarias JC;

XX WPI; 2000-602195/57.

DR Fluorescent apophytochrome-bilin conjugates, e.g. Cph2 (derived from
PT Synecocystis species) and phycoerythrobilin conjugate, useful as
PT fluorescent markers for biological research -

XX Example 1; Page 50; 52pp; English.

XX The present invention is related to fluorescent apophytochrome-bilin
CC conjugates, known as phytofluors. An apoprotein known as Cph2 from
CC Synecocystis species is used as the apophytochrome and the bilin is
CC preferably phycoerythrobilin. The phytofluors are useful as fluorescent
CC markers for biological research. The phytofluors have a long wavelength
CC absorption maxima, a high molar absorption coefficient and the
CC recombinant apoproteins can spontaneously assemble with a variety of
CC bilin chromophore precursors. The present sequence is a phytochrome

Db 4219 lgavvlpqs 4228

RESULT 22

AAW55323
ID AAW55323 standard; peptide; 36 AA.

AC AAW55323;

XX 23-MAR-2001 (first entry)

XX Gene #7 associated peptide #8.

XX Secreted protein; gene therapy; vaccine; cancer; leukemia;
KW autoimmune disease; allergy; inflammation; graft rejection;
KW hyperproliferation; cardiovascular; infection.

XX Homo sapiens.

XX WO200075375-A1.

XX 14-DEC-2000.

XX 02-JUN-2000; 2000WO-US15187.

XX 07-JUN-1999; 99US-0137725.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Birse CE, Duan RD, Soppet DR, Rosen CA, Shi Y;
PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Ni J, Young PE;

XX WPI; 2001-061741/07.

XX Nucleic acids encoding 26 human secreted polypeptides, useful e.g. for
PT preventing, diagnosing and/or treating cancers and for promoting wound
PT healing -

XX Disclosure; Page 29; 530pp; English.

XX The present invention relates to 26 secreted human proteins. The
CC proteins may be used in the prevention, diagnosis and treatment of
CC diseases associated with inappropriate polypeptide expression.
CC For example, they may be used in gene therapy or in vaccines.
CC Typical of diseases which are potentially treatable are cancers
CC (including leukemia), autoimmune diseases, allergies, inflammation,
CC graft rejection, hyperproliferation, cardiovascular diseases
CC (particularly critical limb ischemia and coronary disease) and any
CC involving abnormal angiogenesis, neurodegeneration and/or
CC infectious diseases.

XX Sequence 36 AA;

Query Match 64.4%; Score 29; DB 22; Length 36;
Best Local Similarity 55.6%; Pred. No. 23;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10

DB 9 gsvirlsqa 17
|||::|||

RESULT 23

AAW55323
ID AAW55323 standard; Protein; 155 AA.

AC AAW55323;

XX 15-JUN-1998 (first entry)

XX H. pylori ORF hp2ell1858orf6 protein.

XX

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.

XX Helicobacter pylori.

XX WO9737044-A1.

XX 09-OCT-1997.

XX 27-MAR-1997; 97WO-US05223.

XX 06-DEC-1996; 96US-0761318.

XX 29-MAR-1996; 96US-0625811.

XX 02-APR-1996; 96US-0758731.

XX 25-OCT-1996; 96US-0736905.

XX 28-OCT-1996; 96US-0738859.

XX (ASTR) ASTRA AB.

XX Alm RA, Smith D;

XX WPI; 1997-503122/46.

XX N-PSDB; AAV24732.

XX Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection

XX Claim 14; Page 550; 1145pp; English.

XX This sequence is a H. pylori protein of unspecified function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors. The
CC DNA and probes derived from it may be used for the identification of
CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
CC acid sequences complementary to the DNA act as antisense sequences and
CC can be used to prevent the translation of H. pylori mRNA. Antibodies
CC against the protein can be used in immunoassays to evaluate the abundance
CC and distribution of H. pylori-specific antigens. The genomic sequence of
CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were analysed
CC for ORF of at least 180 nucleotides, and the predicted coding regions
CC defined by computer evaluation. To identify likely H. pylori antigens for
CC vaccine development, the amino acid sequences predicted from various ORF
CC were analysed for significant homology to other known or exported
CC membrane proteins. Having identified and determined the sequences of
CC interest, particular regions can be isolated from H. pylori by PCR
CC amplification for recombinant polypeptide production, e.g. in E. coli
CC hosts.

XX Sequence 155 AA;

Query Match 64.4%; Score 29; DB 18; Length 155;
Best Local Similarity 55.6%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9

DB 101 igsvvldq 109
:||:|:|

RESULT 24

AAW60166
ID AAW60166 standard; Protein; 160 AA.

XX AAW60166;

XX 31-JAN-2000 (first entry)

XX Human endometrium tumour EST encoded protein 226.

XX 16-MAR-1996 (first entry)
XX Erbb-2-specific scFv(FRP)5.
XX Single chain antibody; scFv; antibody engineering; antitumour;
KW tumour antigen binding; T-cell receptor; cytotoxic T-lymphocyte;
KW cell targeting; monoclonal antibody; erbb-2; cancer;
KW adoptive immunotherapy.
XX Synthetic.
XX Key Location/Qualifiers
FH Region 2..30
FT /label= FR1
FT Region 31..35
FT /label= CDR1H
FT Region 36..49
FT /label= FR2
FT Region 50..66
FT /label= CDR2H
FT Region 67..98
FT /label= FR3
FT Region 99..108
FT /label= CDR3H
FT Region 110..120
FT /label= FR4
FT Peptide 121..134
FT /label= Spacer_peptide
FT Region 135..157
FT /label= FR6
FT Region 158..168
FT /label= CDR1L
FT Region 169..183
FT /label= FR7
FT Region 184..190
FT /label= CDR3L
FT Region 191..222
FT /label= FR8
FT Region 223..231
FT /label= CDR3L
FT Region 232..240
FT /label= FR9
XX WO9530014-A1.
XX 09-NOV-1995.
XX 20-APR-1995; 95WO-EP01494.
XX 02-MAY-1994; 94EP-0810244.
XX (CIBA) CIBA GEIGY AG.
XX Groner B, Moritz D;
XX WPI; 1995-393085/50.
XX N-PSDB; AAT05781.
XX New bifunctional proteins for use in killing tumour cells - contg. a
PT tumour antigen binding domain, a hinge region and a zeta chain
PT derived from a T-cell antigen receptor
XX
XX Disclosure; Page 23-25; 46pp; English.
XX Single chain antibody scFv(FRP)5 (AAR85494) is specific for the
CC extracellular domain of erbb-2. The heavy and light chain variable
CC regions of the scfv are derived from mouse monoclonal antibody FRP5
CC (EACACC 90112115), while the framework regions are of human origin.
CC The scfv is used as the tumour antigen binding domain of a
CC bifunctional protein also contg. a hinge region and a zeta chain of a
CC T-cell receptor that is expressed in host cells, esp. cytotoxic
CC T-lymphocytes (CTLs). The CTLs are provided with a defined tumour

CC cell specificity enabling targeting and destruction of the tumour
CC cells.
XX
SQ Sequence 240 AA;
Query Match 64.4%; Score 29; DB 16; Length 240;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 GSIVQLSQS 10
Db 133 gsdqltqs 141
||:||:||
RESULT 27
AAR85495
ID AAR85495 standard; Protein; 240 AA.
XX
AC AAR85495;
XX
DT 16-MAR-1996 (first entry)
XX ScFv(FWP51).
XX Single chain antibody; scFv; antibody engineering; antitumour;
KW tumour antigen binding; T-cell receptor; cytotoxic T-lymphocyte;
KW cell targeting; monoclonal antibody; cancer;
KW adoptive immunotherapy.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Region 2..30
FT /label= FR1
FT Region 31..35
FT /label= CDR1H
FT Region 36..49
FT /label= FR2
FT Region 50..66
FT /label= CDR2H
FT Region 67..98
FT /label= FR3
FT Region 99..109
FT /label= CDR3H
FT Region 110..120
FT /label= FR4
FT Peptide 121..135
FT /label= Spacer_peptide
FT Region 136..158
FT /label= FR6
FT Region 159..164
FT /label= CDR1L
FT Region 170..184
FT /label= FR7
FT Region 185..191
FT /label= CDR3L
FT Region 192..223
FT /label= FR8
FT Region 224..231
FT /label= CDR3L
FT Region 232..240
FT /label= FR9
XX WO9530014-A1.
XX 09-NOV-1995.
XX 20-APR-1995; 95WO-EP01494.
XX 02-MAY-1994; 94EP-0810244.
XX (CIBA) CIBA GEIGY AG.
PA

```

XX  Groner B, Moritz D;
PI  WPI; 1995-393085/50.
DR  N-PSDB; AAT05782.
XX  New bifunctional proteins for use in killing tumour cells - contg. a
XX  antigen in the hinge region and a zeta chain
XX  derived from a T-cell antigen receptor
PI  Disclosure; Page 26-27; 46pp; English.
PS  Single chain antibody scFv(FWP51) (AAR85495) has heavy and light chain
CC  variable regions derived from mouse monoclonal antibody FRP5. The
CC  scFv(225) is a single chain antibody in which the scFv is
CC  used as the tumour antigen binding domain of a bifunctional protein
CC  also contg. a hinge region and a zeta chain of a T-cell receptor that
CC  is expressed in host cells, esp. cytotoxic T-lymphocytes (CTLs). The
CC  CTLs are provided with a defined tumour cell specificity enabling
CC  targeting and destruction of the tumour cells.
XX  Sequence 240 AA;
SQ

```

Query Match 64.4%; Score 29; DB 16; Length 240;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSOS 10
 ||:|:|:|
 Db 134 gsdqltqgs 142

RESULT 28
 AAW05133
 ID AAW05133 standard; Protein; 240 AA.
 XX AAW05133;
 AC AAW05133;
 DT 29-JAN-1997 (first entry)
 DE Single chain antibody scFv(225).
 XX Single chain antibody scFv(225).
 KW Single chain antibody; scFv; monoclonal antibody; Mab; EGF;
 KW epidermal growth factor; receptor; antitumour; cancer; therapy.
 OS Chimeric Mus sp.;
 OS Chimeric synthetic.
 XX Key Location/Qualifiers
 FT Region 1..119
 FT /label= VH_region
 FT /note= "monoclonal antibody 225 VH"
 FT Peptide 120..133
 FT /label= Linker
 FT /note= "synthetic spacer peptide"
 FT Region 134..240
 FT /label= VL
 FT /note= "monoclonal antibody 225 VL"
 XX EP739984-A1.
 XX 30-OCT-1996.
 XX 26-APR-1995; 95EP-0106275.
 XX 26-APR-1995; 95EP-0106275.
 XX (SANT-) SAN TUMORFORSCHUNGS GMBH.
 XX Groner B, Schmidt M, Wels W;
 XX WPI; 1996-478748/48.
 XX N-PSDB; AAT42034.
 XX Bivalent fusion proteins that bind epidermal growth factor receptor
 PT or analogues - and comprise at least two different cell surface

```

DR  N-PSDB; AAT42033.
XX Bivalent fusion proteins that bind epidermal growth factor receptor
XX or analogues - and comprise at least two different cell surface
XX binding domain(s), useful for tumour therapy
XX Example 2; Page 17; 52pp; English.
XX scFv(225) (AAW05133) comprises the single-chain binding region of
CC murine monoclonal antibody 225, which is specific for the human
CC epidermal growth factor receptor. It is encoded by plasmid
CC pMW152-225 (see also AAT42033), constructed by cloning Mab 225
CC VH and VL region cDNAs into a pMT100 vector (pMT100-35) and
CC an effector e.g. cytotoxin, can be produced in bacterial
CC host cells, and are useful as antitumour agents.
XX Sequence 240 AA;
SQ

```

Query Match 64.4%; Score 29; DB 17; Length 240;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSOS 10
 ||:|:|:|
 Db 133 gsdqltqgs 141

RESULT 29
 AAW05134
 ID AAW05134 standard; Protein; 240 AA.
 XX AAW05134;
 AC AAW05134;
 DT 29-JAN-1997 (first entry)
 DE Single chain antibody scFv(FRP5).
 XX Single chain antibody; scFv; monoclonal antibody; Mab; EGF; erbB-2;
 KW epidermal growth factor; receptor; antitumour; cancer; therapy.
 OS Chimeric Mus sp.;
 OS Chimeric synthetic.
 XX Key Location/Qualifiers
 FT Region 1..119
 FT /label= VH_region
 FT /note= "monoclonal antibody FRP5 VH"
 FT Peptide 120..133
 FT /label= Linker
 FT /note= "synthetic spacer peptide"
 FT Region 134..240
 FT /label= VL
 FT /note= "monoclonal antibody FRP5 VL"
 XX EP739984-A1.
 XX 30-OCT-1996.
 XX 26-APR-1995; 95EP-0106275.
 XX 26-APR-1995; 95EP-0106275.
 XX (SANT-) SAN TUMORFORSCHUNGS GMBH.
 XX Groner B, Schmidt M, Wels W;
 XX WPI; 1996-478748/48.
 XX N-PSDB; AAT42034.
 XX Bivalent fusion proteins that bind epidermal growth factor receptor
 PT or analogues - and comprise at least two different cell surface

PT binding domain(s), useful for tumour therapy

XX
PS
XX Example 3; Page 18; 52pp; English.

CC scFv(FRP5) (AAW05134) comprises the single-chain binding region of
CC murine monoclonal antibody FRP5, which is specific for the human
CC epidermal growth factor receptor erbB-2 (Neu/HER-2). It is encoded
CC by plasmid pMW152-5 (see also AAT42034). Novel bivalent proteins
CC (see also AAW05135-44), some of them including scFv(FRP5) and an
CC effector, e.g. cytotoxin, molecule can be produced in bacterial
CC host cells, and are useful as antitumour agents.

XX
SQ Sequence 240 AA;

Query Match 64.4%; Score 29; DB 17; Length 240;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVOLSQS 10

Db 133 gsdqitqqs 141

RESULT 30

AAW25784
ID AAW25784 standard; Protein; 240 AA.

XX
AC AAW25784;

XX
DT 22-DEC-1997 (first entry)

XX
DE Anti-gpl30 antibody derived scFv.

XX
KW Single chain Fv; scFv; antibody; E. coli; anti-T3 antibody;
KW anti-gpl30 antibody; inclusion body; chaperonin.

XX
OS Synthetic.

XX
PN JP09220092-A.

XX
PD 26-AUG-1997.

XX
PF 15-FEB-1996; 96JP-0027622.

XX
PR 15-FEB-1996; 96JP-0027622.

XX
PA (TOYJ) TOSOH CORP.

DR
WPI; 1997-474306/44.

DR
N-PSDB; AAW91615.

XX
PT Producing single chain Fv antibody in Escherichia coli - by
PT expression in an inclusion body, followed by protein folding or by
PT co-expression with a chaperonin as a soluble fraction

XX
PS Claim 4; Page 6-7; 9pp; Japanese.

XX
CC The sequences given in AAW25783-84 represent single chain Fv (scFv)
CC antibodies which are produced in E.coli. The scFv's are derived
CC from an anti-T3 antibody or an anti-gpl30 antibody and are produced
CC either by: (1) expression as an inclusion body, followed by folding
CC (i.e. by denaturation and solubilisation) or (2) expression as a
CC soluble fraction by co-expression with chaperonin. Using the methods,
CC scFv can be produced in E. coli, either in a soluble fraction or in
CC inclusion bodies. Like the parent double chain antibodies, the scFv
CC polypeptides are useful in immunological diagnosis or for separation
CC and purification, but they lessen the side effects caused by antibody
CC constant regions. The availability of scFv polypeptides contributes
CC to fundamental research and development of diagnostic and therapeutic
CC drugs.

XX
SQ Sequence 240 AA;

Query Match 64.4%; Score 29; DB 18; Length 240;
Best Local Similarity 77.8%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSIVOLSQS 10

Db 119 gsdvqlqqs 127

RESULT 31

AAW71243

ID AAW71243 standard; Protein; 240 AA.

XX
AC AAW71243;

XX
DT 20-NOV-1998 (first entry)

XX
DE scFv comprising heavy and light chain variable domains of antibody 14E1.

XX
KW Antibody 14E1; human epidermal growth factor receptor; EGF-R; treatment;
KW carcinoma; EGF-RvIII; in vitro purging; breast carcinoma.

XX
OS Mus sp.

XX
FH Key Location/Qualifiers

FT Protein 1..119

FT Peptide /note= "Heavy chain variable domain"

FT Protein 120..134

FT Peptide /note= "synthetic linker"

FT Protein 135..240

FT Protein /note= "light chain variable domain"

XX
PN WO9836074-A2.

XX
PD 20-AUG-1998.

XX
PF 13-FEB-1998; 98WO-IB00413.

XX
PR 13-FEB-1997; 97US-0800198.

XX
PA (SCHD) SCHERING AG.

XX
PI Schmidt M, Schneider DW, Vakalopoulou E, Wells WS;

XX
DR WPI; 1998-457176/40.

XX
DR N-PSDB; AAV54790.

XX
PT New single chain polypeptide that binds to epidermal growth factor
PT receptor - derived from antibody 14E1, and related nucleic acid and
PT double-chain forms, used for treatment and diagnosis of cancer or
PT other cell proliferative diseases

XX
PS Claim 11; Fig 12; 48pp; English.

XX
CC The present sequence represents a single-chain polypeptide (scFv) of
CC the invention. It is derived from murine monoclonal antibody 14E1. This
CC antibody is active against the human epidermal growth factor receptor
CC (EGF-R). The heavy and light chains of antibody 14E1 are used to
CC construct a scFv that has binding affinity for an EGF-R. The scFv
CC comprises two segments which contain the binding parts of the heavy and
CC light chain variable domains of the monoclonal antibody 14E1. The scFvs
CC of the invention are used to treat carcinomas (gliomas, melanomas and
CC tumours) that overexpress EGF-R, specifically EGF-RvIII, relative to
CC normal cells, particularly by blocking proliferation of such cells by
CC preventing receptor activation. They can also be used for in vitro
CC purging. When the scFv is appropriately labelled, it can be used
CC diagnostically to image cells that express EGF-R, specifically breast
CC carcinoma.

XX
SQ Sequence 240 AA;

Query Match 64.4%; Score 29; DB 19; Length 240;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSVOLQS 10
 II :II:II
 DB 133 gsdilqtqs 141

RESULT 32
 AAR26979
 ID AAR26979 standard; Protein; 241 AA.

XX AAR26979;
 XX 11-FEB-1993 (first entry)

XX Fv heavy chain/light chain variable domain fusion protein.
 XX Minigene; pW18/1; monoclonal antibody; kappa; light chain;
 KW variable region; pW31; tumour.
 XX Mus musculus.

XX Key Location/Qualifiers
 FH Domain 2..120
 FT Label= FRP5_heavy_chain_variable_domain
 FT Region 32..36
 FT Label= CDR1H
 FT Region 51..67
 FT Label= CDR2H
 FT Region 100..109
 FT Label= CDR3H
 FT Peptide /label= Linker
 FT Domain 136..241
 FT Label= FRP5_light_chain_variable_domain
 FT Region 159..169
 FT Label= CDR1L
 FT Region 185..191
 FT Label= CDR2L
 FT Region 225..232
 FT Label= CDR3L

PN EP502812-A.

XX 09-SEP-1992.

XX 27-JAN-1992; 92EP-0810056.

XX 05-FEB-1991; 91EP-0810079.

XX (CIBA) CIBA GEIGY AG.

XX Groner B, Hardman N, Harwerth I, Hynes NE, Wells WS;
 Zwickl M;

XX WPI; 1992-302096/37.

XX N-PSDB; AAQ28256.

XX Recombinant antibodies directed to growth factor receptor C-erbB-2 -
 for diagnosing and treating tumours expressing C-erbB-2 e.g. breast
 or ovarian tumours

XX Disclosure; Page 31-33; 67pp; English.

XX The sequence given is the Fv heavy chain/light chain variable domain
 fusion protein which binds to the extracellular domain of the growth
 factor receptor C-erbB-2. This recombinant antibody can be used for
 the qualitative and quantitative determination of C-erbB-2. This can
 be used for monitoring or in-vivo localisation of tumours
 overexpressing C-erbB-2.

CC

XX SQ Sequence 241 AA;
 Query Match 64.4%; Score 29; DB 13; Length 241;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSVOLQS 10
 II :II:II
 DB 134 gsdilqtqs 142

RESULT 33
 AAR26981

ID AAR26981 standard; Protein; 241 AA.

XX AAR26981;

XX 11-FEB-1993 (first entry)

XX FWP51 fusion protein.

XX Monoclonal antibody; light chain; heavy chain; tumour; c-erbB-2;
 KW variable region.

XX Synthetic.

XX Key Location/Qualifiers
 FH Domain 2..121
 FT Label= FWP51_heavy_chain_variable_domain
 FT Region 32..36
 FT Label= CDR1H
 FT Region 51..67
 FT Label= CDR2H
 FT Region 100..110
 FT Peptide /label= CDR3H
 FT Domain 122..136
 FT Label= Synthetic_spacer
 FT Region 137..241 FWP52_light_chain_variable_domain
 FT Label= CDR1L
 FT Region 166..170
 FT Label= CDR2L
 FT Region 186..192
 FT Label= CDR3L
 FT Region 225..232

PN EP502812-A.

XX 09-SEP-1992.

XX 27-JAN-1992; 92EP-0810056.

XX 05-FEB-1991; 91EP-0810079.

XX (CIBA) CIBA GEIGY AG.

XX Groner B, Hardman N, Harwerth I, Hynes NE, Wells WS;
 Zwickl M;

XX WPI; 1992-302096/37.

XX N-PSDB; AAQ28260.

XX Recombinant antibodies directed to growth factor receptor C-erbB-2 -
 for diagnosing and treating tumours expressing C-erbB-2 e.g. breast
 or ovarian tumours

XX Disclosure; Page 43-45; 67pp; English.

XX The sequence given is encoded by the single-chain Fv fusion gene
 comprising the monoclonal antibody FWP51 heavy and kappa light chain
 variable domains (see AAQ28258-9). This recombinant antibody can be used

CC

CC for the qualitative and quantitative determination of c-erbB-2. This
 CC can be used for monitoring or in-vivo localisation of tumours
 CC overexpressing c-erbB-2.

XX SQ Sequence 241 AA;
 Query Match 64.4%; Score 29; DB 13; Length 241;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
 II :||:|
 Db 135 gsdqltqs 143

RESULT 34
 AAB31426
 ID AAB31426 standard; protein; 242 AA.

XX AC AAB31426;
 XX DT 20-APR-2001 (first entry)

XX DE Protein used for immunising against shed antigen-specific B cells.
 XX KW Vaccine; shed antigen-specific B cell; idiotype antibody;
 KW immune complex-mediated disease; autoimmune disease;
 KW humoral immune response; cancer.

XX OS Synthetic.
 XX PN WO200076319-A1.

XX PD 21-DEC-2000.
 XX PF 16-JUN-2000; 2000WO-US16677.

XX PR 16-JUN-1999; 99US-0139521.
 XX PR 15-JUN-2000; 2000US-0594985.

XX PA (BIOC-) BIOCRYSTAL LTD.

XX PI Nelson MB, Barbera-Guillem E;

XX DR WPI; 2001-080635/09.

XX PT Inducing an immune response against shed antigen-specific B cell
 PT idiotypes, by administering a vaccine formulation comprising
 PT polynucleotides encoding an idiotype determinant or peptides
 PT comprising an idiotype determinant

XX PS Example 2; Page 71; 81pp; English.

XX CC The present sequence represents a synthetic protein which may be used
 CC in vaccines for inducing immune responses of the invention. The protein
 CC comprises a VH region linked to a VL region. The specification describes
 CC a method for inducing an immune response reactive with idiotypes present
 CC on shed antigen-specific B cells (SAB) of an individual. The method
 CC involves administering a vaccine formulation comprising polynucleotide
 CC encoding an idiotype of an antibody that binds to an epitope of shed
 CC antigen. The method is useful for inducing an immune response reactive
 CC with idiotypes present on SAB of an individual. The method is useful for
 CC depleting shed antigen-specific B cells, and for treating immune
 CC complex-mediated disease progression in organ specific autoimmune disease
 CC exacerbated by humoral immune response against groups expressed on shed
 CC antigen, or by plasma cell production of antibodies against groups of
 CC shed antigen. It is useful in cancer therapy and for treating autoimmune
 CC disease.

XX SQ Sequence 242 AA;

Query Match 64.4%; Score 29; DB 22; Length 242;
 Best Local Similarity 77.8%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSIVQLSQS 10
 II |||:|
 Db 123 gsqvqlqqs 131

RESULT 35
 AAR77606
 ID AAR77606 standard; Protein; 248 AA.

XX AC AAR77606;

XX DT 15-MAR-1996 (first entry)

XX DE Murine 5G1.1M1 scFv.

XX KW Complement C5; haemolysis; kidney; glomerulonephritis;
 KW monoclonal antibody; antiinflammatory; antibody engineering;
 KW humanised antibody; complementarity determining region; CDR
 KW single chain antibody; scfv.

XX OS Mus sp.

XX FH Key Location/Qualifiers
 FT Region 28..34
 FT /label= CDR-L1
 FT Region 52..54
 FT /label= CDR-L2
 FT Region 93..98
 FT /label= CDR-L3
 FT Region 156..159
 FT /label= CDR-H1
 FT Region 179..183
 FT /label= CDR-H2
 FT Region 226..236
 FT /label= CDR-H3

XX PN WO9529697-A1.

XX PD 09-NOV-1995.

XX PF 01-MAY-1995; 95WO-US05688.

XX PR 02-MAY-1994; 94US-0236208.

XX PA (ALEX-) ALEXION PHARM INC.

XX PI Evans MJ, Matis L, Mueller EE, Nye SH, Rollins S;
 PI Rother RP, Springhorn J P, Squinto SP, Thomas TC;
 PI Wang Y, Wilkins JA;

XX DR WPI; 1995-392923/50.
 XX DR N-PSDB; AAT08479.

XX PT Treating glomerulonephritis with antibody against complement C5
 PT component - to inhibit complement induced cell lysis

XX PS Example 11; Page 107-110; 181pp; English.

XX CC Murine scFv 5G1.1M1 (AAR77606) is the product of a DNA construct
 CC (AAT08479) obt'd. from DNA of hybridoma ATCC HB 11625, which produces
 CC anti-C5 monoclonal antibody (Mab) 5G1.1. Recombinant scFv 5G1.1M1 is
 CC obt'd. by expression of this DNA in Escherichia coli using vector pET
 CC Trc SOS/NI. The light and/or heavy chain CDRs of scFv 5G1.1M1 can
 CC be combined with CDRs from other 5G1.1-derived antibodies, Fds and
 CC light chains (AAR77607-16) in the prodn. of recombinant, including
 CC humanised, antibodies that retain the ability of Mab 5G1.1 to block
 CC human complement C5a generation and thus to reduce glomerular
 CC inflammation and kidney dysfunction associated with
 CC glomerulonephritis.


```

Db 135 gsdqltqs 143
|| :||:|
RESULT 38
AAW95667
ID AAW95667 standard; protein; 248 AA.
AC AAW95667;
XX
DT 08-JUN-1999 (first entry)
XX
DE Mus musculus anti-IgE e26 sfv fragment.
XX
KW sfv fragment; IgE; antibody; anti-IgE; reduction; prevention;
KW histamine; production; hypersensitivity; allergen; anaphylaxis;
KW atopic allergy; asthma; allergic rhinitis; conjunctivitis; hay fever;
KW eczema; anaphylactic shock; urticaria.
XX
OS Mus musculus.
XX
PN WO9901556-A2.
XX
PD 14-JAN-1999.
XX
PF 30-JUN-1998; 98WO-US13410.
XX
PR 02-JUL-1997; 97US-0887352.
XX
PA (GETH ) GENENTECH INC.
PI Jardieu PM, Lowe J, Lowman HB, Presta LG;
XX
DR WPI; 1999-106057/09.
XX
PT Improving affinity of polypeptides, particularly anti-IgE antibodies
PT - by identifying aspartyl residues which undergo isomerisation and
PT substituting alternative residues and screening for affinity against
PT the target
XX
PS Disclosure; Pages 102-103; 129pp; English.
XX
CC The sequence is that of the e26 sfv fragment. It
CC was used as part of a method to improve the affinity of anti-IgE
CC antibodies such as e26 and e27. The e26 and e27 antibodies can
CC be used for reducing or preventing IgE mediated production of
CC histamine in a mammal. They can be used for treating a disorder
CC mediated by IgE such as hypersensitivity, atopic allergy, asthma,
CC allergic rhinitis, conjunctivitis, hay fever, eczema, anaphylactic
CC shock and urticaria. The antibodies can also be used for affinity
CC purification, detection and diagnosis.
XX
SQ Sequence 248 AA;

Query Match 64.4%; Score 29; DB 20; Length 248;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
|| :||:|
Db 135 gsdqltqs 143

RESULT 39
AAB76956
ID AAB76956 standard; protein; 248 AA.
XX
AC AAB76956;
XX
DT 17-APR-2001 (first entry)
XX
DE SFv fragment of e26 SEQ ID 22.

```

```

XX Antibody; antiasthmatic; antiallergic; ophthalmological; dermatological;
XX antiinflammatory; Ig E; immunoglobulin E; asthma; allergic rhinitis;
XX conjunctivitis; eczema; urticaria; food allergy.
XX Synthetic.
XX OS
XX PN US6172213-B1.
XX
XX PD 09-JAN-2001.
XX
XX PF 30-JUN-1998; 98US-0109207.
XX
XX PR 02-JUL-1997; 97US-0051554.
XX
XX (GETH ) GENENTECH INC.
XX
XX Lowman HB, Presta LG, Jardieu PM, Lowe J;
PI WPI; 2001-122353/13.
DR
XX
XX New nucleic acid encoding anti-immunoglobulin E antibody with improved
XX properties, produced by substituting aspartyl residues in unimproved
XX immunoglobulin E prone to isomerization by other residues by affinity
XX maturation with phage display
XX
XX Disclosure; Fig 14; 87pp; English.
XX
CC This invention relates to a nucleotide sequence encoding an antibody
CC with improved anti-IgE antibody activity. The antibody has improved
CC action due to a process comprising, a) identifying aspartyl residues
CC prone to isomerisation in unimproved anti-IgE (immunoglobulin E)
CC antibody; b) substituting alternative residues to create candidate
CC molecules, and c) selecting those candidate molecules which display
CC affinity against the target molecule. Use of the antibody results in
CC antiasthmatic; antiallergic; ophthalmological; dermatological and
CC antiinflammatory activity. The antibodies are useful for treating
CC IgE-mediated disorders such as asthma, allergic rhinitis, conjunctivitis,
CC eczema, urticaria and food allergies. The mutant antibodies produced by
CC the above mentioned nucleic acids may also be used as affinity
CC purification agents and in diagnostic assays for detecting the expression
CC of an antigen of interest in specific cell, tissues or serum. Amino acid
CC sequences AAB76936-AAB76960 represent fragments of anti-IgE antibodies of
CC the invention. Polynucleotide sequence AAF69253 represents an expression
CC plasmid used in the course of the invention, and oligonucleotides
CC AAF69254 - AAF69271 are used in the generation of affinity improved
CC anti-IgE antibodies.
XX
SQ Sequence 248 AA;

Query Match 64.4%; Score 29; DB 22; Length 248;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
|| :||:|
Db 135 gsdqltqs 143

RESULT 40
AAB76957
ID AAB76957 standard; protein; 248 AA.
XX
AC AAB76957;
XX
DT 17-APR-2001 (first entry)
XX
DE SFv fragment of e27 SEQ ID 23.
XX
KW Antibody; antiasthmatic; antiallergic; ophthalmological; dermatological;
KW antiinflammatory; Ig E; immunoglobulin E; asthma; allergic rhinitis;
KW conjunctivitis; eczema; urticaria; food allergy.

```

XX Synthetic.
 OS US6172213-B1.
 PN XX
 PD XX
 PP 09-JAN-2001.
 PF 30-JUN-1998; 98US-0109207.
 PR 02-JUL-1997; 97US-0051554.
 XX (GETH) GENENTECH INC.
 PA Lowman HB, Presta LG, Jardieu PM, Lowe J;
 PI WPI: 2001-122353/13.
 XX
 XX New nucleic acid encoding anti-immunoglobulin E antibody with improved
 PT properties, produced by substituting aspartyl residues in unimproved
 PT immunoglobulin E prone to isomerization by other residues by affinity
 PT maturation with phage display -
 XX
 XX Claim 3; Fig 14; 87pp; English.
 CC This invention relates to a nucleotide sequence encoding an antibody
 CC with improved anti-IgE antibody activity. The antibody has improved
 CC action due to a process comprising, a) identifying aspartyl residues
 CC prone to isomerization in unimproved anti-IgE (immunoglobulin E)
 CC antibody; b) substituting alternative residues to create a candidate
 CC antibody against the target molecule. Use of the antibody results in
 CC affinity against the target molecule. Use of the antibody results in
 CC antiasthmatic; antiallergic; ophthalmological; dermatological and
 CC antiinflammatory activity. The antibodies are useful for treating
 CC IgE-mediated disorders such as asthma, allergic rhinitis, conjunctivitis,
 CC eczema, urticaria and food allergies. The mutant antibodies produced by
 CC the above mentioned nucleic acids may also be used as affinity
 CC reagents for the detection of an antigen of interest in specific cell, tissues or serum. Amino acid
 CC sequences AAB76936-AAB76960 represent fragments of anti-IgE antibodies of
 CC the invention. Polynucleotide sequence AAF69253 represents an expression
 CC plasmid used in the course of the invention, and oligonucleotides
 CC AAF69254 - AAF69271 are used in the generation of affinity improved
 CC anti-IgE antibodies.
 XX
 XX Sequence 248 AA;
 SQ

Query Match 64.4%; Score 29; DB 22; Length 248;
 Best Local Similarity 66.7%; Pred. No. 2e+02; 1; Indels 0; Gaps 0;
 Matches 6; Conservative 2; Mismatches 1;
 QY 2 GSVOLSQS 10
 II :||:II
 Db 135 gsdqltqs 143

RESULT 41
 ID AAF80152 standard; protein; 251 AA.
 AC AAP80152;
 XX 20-OCT-1990 (first entry)
 XX Multifunctional protein.
 XX Multifunctional protein; biosynthetic antibody binding site;
 KW murine monoclonal 26-10;
 XX WO8809344-A.
 FN 01-DEC-1988.
 XX
 XX

PF 19-MAY-1988; 88WO-US01737.
 XX
 PR 21-MAY-1987; 87US-0052800.
 XX
 PA (CREA-) CREATIVE BIOMOLECULES INC.
 XX
 XX Huston JS, Oppermann H;
 PI WPI: 1988-353928/49.
 DR N-PSDB; AAN80178.
 XX
 XX Recombinant multifunctional protein - having antibody binding site and a
 PT sequence for biological activity, ion sequestering or binding to a solid
 PT support.
 XX
 XX Disclosure; 15pp; English.
 CC The multifunctional protein comprises a single chain biosynthetic Ab
 CC binding site (comprising sequences mimicking the light chain and heavy
 CC chain variable regions of murine monoclonal 26-10) with the specificity
 CC of murine monoclonal 26-10, linked through a spacer to the C-terminus of
 CC protein A, here inserted as a leader, and constituting a binding site for
 CC a specific antigen. The construct contains
 CC Asp-Pro a dilute acid cleavage site). The construct contains
 CC binding sites for both FC and digoxin, and its structure can be
 CC summarised as: FB-Asp-Pro-VL-(Gly4-Ser)3-VH. The protein can be used for
 CC specific binding assays, affinity purification, biocatalysts, targeting,
 CC imaging an immunological treatment of oncogenic and infectious diseases.
 CC In order to leave the target molecule intact, the protein is designed
 CC to have high stability. It reaches target tissues rapidly and is quickly
 CC cleared from the body. It also has reduced immunogenicity and its design
 CC facilitates binding to other moieties in drug targeting and imaging
 CC applications.
 XX
 XX Sequence 251 AA;
 SQ

Query Match 64.4%; Score 29; DB 9; Length 251;
 Best Local Similarity 77.8%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0;
 QY 2 GSVOLSQS 10
 II :||:II
 Db 131 gsevlqqs 139

RESULT 42
 ID AAB28159 standard; protein; 253 AA.
 AC AAB28159;
 XX 08-FEB-2001 (first entry)
 XX Murine anti-EGP-2 single chain Fv fragment.
 XX Murine: immunoglobulin; antigen-binding; framework region; carcinoma;
 KW C-erbB2; carcinoma associated antigen.
 XX Mus sp.
 XX WO2000061635-A2.
 XX 19-OCT-2000.
 XX 10-APR-2000; 2000WO-EP03176.
 XX 09-APR-1999; 99EP-0107030.
 XX (UYZU-) UNIV ZUERICH.
 XX (PLUEK) PLUECKTHUN A.
 XX Plueckthun A, Honegger A, Willuda J;
 PI

XX WPI; 2000-679468/66.
 XX Stabilizing chimeric immunoglobulin (Ig) involves setting up a
 PT stabilizing antigen binding Ig or its fragment by replacing one or more
 PT residues present in acceptor Ig by those residues present in donor Ig
 PT
 XX
 XX Claim 6; Page 51; 51pp; English.
 PS
 PS The present invention relates to a method for stabilising a chimeric
 CC immunoglobulin (Ig). The method comprises identifying antigen-binding
 CC groups derived from donor Ig and framework regions derived from an
 CC acceptor Ig. The present sequence is one such donor Ig fragment.
 CC One or more of the residues present at the positions in the present
 CC sequence are replaced by those present at the corresponding positions in
 CC the donor Ig, after comparing the structural features of the VH domains
 CC of the acceptor Ig and the donor Ig. The acceptor Ig fragment used in
 CC the present invention is human anti-c-ErbB2 single chain Fv fragment 4D5
 CC (AAB28158). The method of the present invention is useful for producing
 CC a pharmaceutical composition which can be used for treating human
 CC carcinomas, since c-erbB2 is a carcinoma-associated antigen.
 XX
 XX Sequence 253 AA;
 SQ

Query Match 64.4%; Score 29; DB 21; Length 253;
 Best Local Similarity 77.8%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
 DB 136 gsqvqlqqs 144
 || ||| ||
 || ||| ||

RESULT 43
 AAR20185
 ID AAR20185 standard; peptide; 271 AA.
 XX
 XX AAR20185;
 AC
 DT 15-APR-1992 (first entry)
 XX
 XX Sequence of the shortened hinge version of the B72.3 single
 DE chain Fv hinge.
 DE
 XX Fv fragment; in vivo diagnosis; therapy; antibody.
 KW
 XX Homo sapiens.
 OS
 XX
 XX WO9119739-A.
 PN
 XX 26-DEC-1991.
 PD
 XX
 XX 11-JUN-1991; 91WO-GB00935.
 PF
 XX 11-JUN-1990; 90GB-0012995.
 PR
 XX (CELL-) CELLTech LTD.
 XX
 XX King DJ, Mountain A, Owens RJ, Yarranton GT;
 XX
 XX WPI; 1992-024365/03.
 DR
 DR N-PSDB; AAQ20381.
 XX
 XX New multivalent antigen-binding proteins - comprise Fv fragment
 PT linked to at least 1 other Fv fragment spacer and useful for
 PT in-vivo diagnosis or therapy
 PT
 XX Example; Fig 5; 54pp; English.
 PS
 PS The bivalent antigen-binding protein B72.3 Fv (AAR20184) is an example
 CC of the antigen-binding proteins of the invention. They comprise a

CC first Fv fragment bound to at least one other Fv fragment by a
 CC linker which keeps the Fv fragment apart. The connecting structure
 CC of AAR20184 comprises a joining sequence derived from a human IgG1
 CC domain linked to a complete human IgG4 hinge region. AAR20185 is the
 CC sequence of the shortened hinge version of the B72.3 single chain Fv
 CC hinge.
 XX
 XX Sequence 271 AA;
 SQ

Query Match 64.4%; Score 29; DB 13; Length 271;
 Best Local Similarity 77.8%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
 DB 144 gsqvqlqqs 152
 || ||| ||
 || ||| ||

RESULT 44
 AAR34672
 ID AAR34672 standard; Protein; 272 AA.
 XX
 XX AAR34672;
 AC
 DT 25-AUG-1993 (first entry)
 XX
 XX 26-10 sFv.
 DE
 XX Heavy; light; variable; VH; VL; region; antidigitoxin; monoclonal;
 KW antibody; Mab; 26-10; Fv; fragment; antigen; binding site; linker;
 KW expression-secretion system; T7 promoter; signal peptide; molecule;
 KW polymerase chain reaction; PCR; single chain.
 XX
 XX Synthetic.
 OS
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..25
 FT /note= "Signal peptide"
 FT Region 26..138
 FT /note= "26-10 VL sequence"
 FT Peptide 139..153
 FT /label= Linker peptide
 FT Region 154..272
 FT /note= "26-10 VH sequence"
 XX
 XX WO9308300-A.
 PN
 XX 29-APR-1993.
 PD
 XX 16-OCT-1992; 92WO-US08881.
 PF
 XX 18-OCT-1991; 91US-0777709.
 PR
 XX (UYCA-) UNIV CALGARY.
 PA
 XX Anthony JG, Ng SC, Wong S;
 PI
 XX WPI; 1993-152491/18.
 DR
 DR N-PSDB; AAQ41069.
 DR
 XX Expression-secretion vectors - for prodn. of biologically active
 PT antibody Fv fragments or single chain Fv molecules
 PT
 XX Disclosure; Fig 6; 51pp; English.
 PS
 XX This sequence represents a single chain Fv molecule which contains
 CC the heavy and light variable portions (VH and VL) of antidigitoxin
 CC monoclonal antibody (MAB) 26-10 (see also AAR34670-71), a signal
 CC sequence (see AAR38527-28) and a linker between the two variable
 CC regions. The DNA sequence encoding this protein was used in the
 CC construction of an expression-secretion system for the production
 CC of biologically active Fv fragments. The system also contains a

CC DNA sequence encoding the T7 promoter. The secretion-expression
 CC construct contains a T7 promoter, a signal sequence, a fusion
 CC comprising only the VH or VL regions and single chain molecules may
 CC be used to specifically bind one or more of the same antigens as the
 CC full length antibody from which they are derived.

XX Sequence 272 AA;

Query Match 64.4%; Score 29; DB 14; Length 272;
 Best Local Similarity 77.8%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSVQLSOS 10
 II IIII II
 Db 152 gsevlqqs 160

RESULT 45
 AAW00557
 ID AAW00557 standard; Protein: 272 AA.
 XX
 AC AAW00557;
 XX
 DT 31-OCT-1996 (first entry)
 XX
 DE Nematode salivary protein antibody construct.
 XX
 KW Cyst nematode; salivary protein; svp49; transgenic plant;
 KW Cyst nematode; salivary protein; svp49; transgenic plant;
 KW Heterodera; Globodera; Punctodera; monoclonal antibody.
 XX
 OS Mus musculus.

XX Key Location/Qualifiers
 FT Domain 1..112
 FT Region 25..39 Light_chain_variable_domain
 FT Region 55..61 /label= CDR
 FT Region 94..102 /label= CDR
 FT Region 119..127 CDR
 FT Peptide /label= Gly-Ser-linker
 FT Domain 128..272
 FT Region 158..162 /label= Heavy_chain_variable_domain
 FT Region 177..193 CDR
 FT Region 226..232 /label= CDR
 FT Region 226..232 /label= CDR

XX WO9622372-A2.
 XX 25-JUL-1996.
 XX 17-JAN-1996; 96WO-NL00033.
 XX 17-JAN-1995; 95EP-0200110.
 XX (RIJK-) RIJSLANDBOUW Hogeschool Wageningen.
 XX Bakker J, Schots A, Stiekema WJ;
 XX WPI; 1996-354529/35.
 XX N-PSDB; AAT35691.
 XX Antibody against stage two juvenile nematode salivary proteins
 XX used in the production of transgenic plants, resistant to nematode
 XX infection

PS Claim 7; Page 17-19; 25pp; English.

XX An engineered antibody construct (AAW00557) comprises the 2 variable
 CC regions of an antibody that specifically reacts with a salivary
 CC protein of cyst nematodes, fused by a 15-mer linker peptide. The
 CC antibody is pref. mouse monoclonal antibody, AAW00557, AAW00558
 CC and svp49 (see also AAW00556) of cyst nematode second stage juveniles
 CC (J2) and which was raised using J2 protein fractions of Globodera
 CC rostochiensis. Transgenic plants expressing such antibody
 CC constructs are protected against attack from Heterodera, Punctodera
 CC and Globodera cyst nematodes.

XX Sequence 272 AA;

Query Match 64.4%; Score 29; DB 17; Length 272;
 Best Local Similarity 77.8%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 GSVQLSOS 10
 II IIII II
 Db 126 gsevlqqs 134

Search completed: June 28, 2001, 11:54:19
 Job time: 80 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: June 28, 2001, 11:57:04 ; Search time 42.81 Seconds
(without alignments)
17.794 Million cell updates/sec

Title: US-09-439-313-573
Perfect score: 45
Sequence: 1 MGSIVQLSQS 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	33	73.3	3413	2 T17467	rifamycin polyketide synthase modules 9-10 - Anycolatopsis mediterranei
2	31	68.9	304	2 T03597	phosphoprotein phosphatase
3	31	68.9	394	2 T24860	hypothetical protein
4	31	68.9	436	2 D83498	C4-dicarboxylate t
5	31	68.9	468	2 A9289	peroxisome prolifer
6	31	68.9	468	2 J22085	peroxisome prolifer
7	31	68.9	468	2 A5288	peroxisome prolifer
8	31	68.9	468	2 I56603	hypothetical prote
9	31	68.9	617	2 T46319	adenylate cyclase
10	31	68.9	838	2 A38172	adenylate cyclase
11	30	66.7	117	2 S32083	sucrose-phosphate
12	30	66.7	161	2 C75331	transcription elon
13	30	66.7	193	2 S78130	ribosomal protein
14	30	66.7	241	2 C83133	probable short-cha
15	30	66.7	280	2 G81421	probable flagellar
16	30	66.7	305	2 S32237	hypothetical prote
17	30	66.7	364	2 T48752	GRR1 related prote
18	30	66.7	382	2 E85615	probable transport
19	30	66.7	382	2 A64829	membrane protein y
20	30	66.7	425	2 T08946	hypothetical prote
21	30	66.7	503	2 E92699	carboxypeptidase r
22	30	66.7	539	2 B72707	hypothetical prote
23	30	66.7	657	2 S30288	C4-dicarboxylate s
24	30	66.7	662	2 T17339	hypothetical prote
25	30	66.7	750	2 S75636	sensory transducti
26	30	66.7	880	1 B33926	DNA-directed RNA p
27	30	66.7	964	2 A84693	hypothetical prote
28	30	66.7	968	2 F84693	hypothetical prote
29	30	66.7	972	2 E84693	hypothetical prote

30	30	66.7	1059	2 T12195	sucrose-phosphate
31	30	66.7	1277	2 S53043	probable membrane
32	30	66.7	1341	2 T17285	hypothetical prote
33	30	66.7	1520	2 T23620	hypothetical prote
34	30	66.7	2559	2 T09144	probable guanine n
35	30	66.7	5069	2 T17464	rifamycin polyketi
36	29	64.4	188	2 S12393	hypothetical prote
37	29	64.4	225	2 G86748	hypothetical prote
38	29	64.4	254	2 F84145	transposase (23) B
39	29	64.4	261	2 H84149	transposase (27) B
40	29	64.4	274	2 T29574	hypothetical prote
41	29	64.4	285	2 H71938	flagellar motor sw
42	29	64.4	287	2 F64648	fly protein - Hel
43	29	64.4	305	2 A40573	clathrin heavy cha
44	29	64.4	340	2 H70790	hypothetical prote
45	29	64.4	378	2 A85877	probable prophage
46	29	64.4	380	2 T19096	hypothetical prote
47	29	64.4	384	2 S64074	hypothetical prote
48	29	64.4	459	1 S04904	glutamate dehydrog
49	29	64.4	463	2 A65159	servi-tRNA(Sec) se
50	29	64.4	463	2 C86034	hypothetical prote
51	29	64.4	503	1 QX86P	surface glycoprote
52	29	64.4	504	2 T50270	hypothetical prote
53	29	64.4	511	2 S47860	amidophosphoribos
54	29	64.4	540	2 S41864	kinesin light chai
55	29	64.4	563	2 S47997	kinesin light chai
56	29	64.4	573	2 JQ0135	hypothetical 62.8K
57	29	64.4	582	2 T39931	probable transcrip
58	29	64.4	610	2 H83243	probable ATP-bindi
59	29	64.4	622	1 A49344	cell wall assembly
60	29	64.4	679	2 I48084	gibbon ape leukemi
61	29	64.4	958	2 T10679	hypothetical prote
62	29	64.4	1171	2 T17454	diaphanous-related
63	29	64.4	1262	2 T30524	protein phosphatas
64	29	64.4	1569	2 G02757	clathrin heavy cha
65	29	64.4	1640	2 T09522	clathrin heavy cha

ALIGNMENTS

RESULT 1	
T17467	rifamycin polyketide synthase modules 9-10 - Anycolatopsis mediterranei
C:Species:	Anycolatopsis mediterranei
C:Date:	02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C:Accession:	T17467
R:Schupp, T.	
submitted to the EMBL Data Library, December 1997	
A:Reference number:	Z18802
A:Accession:	T17467
A>Status:	preliminary; translated from GB/EMBL/DDBJ
A:Molecule type:	DNA
A:Residues:	1-3413 <SCH>
A:Cross-references:	EMBL:AJ223012; NID:el227119; PID:el227124; PIDN:CAAL1039.1
A:Experimental source:	strain LBG A3136
C:Superfamily:	acyl carrier protein homology
C:Keywords:	carrier protein
F:1608-1679/Domain:	acyl carrier protein homology <ACP1>
F:3334-3405/Domain:	acyl carrier protein homology <ACP2>
Query Match	73.3%; Score 33; DB 2; Length 3413;
Best Local Similarity	60.0%; Pred. No. 1.7e+02;
Matches	6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10

Db 2588 LGAVVQLPQS 2597

RESULT 2
T03597

phosphoprotein phosphatase (EC 3.1.3.16) 1, npp3 - common tobacco
 C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
 C;Accession: T03597
 R;Suhr, M.; Cho, H.; Kim, Y.; Liu, J.; Lee, H.
 Plant Mol. Biol. 36, 315-322, 1998
 A;Title: Multiple genes encoding serine/threonine protein phosphatases and their differ
 A;Reference number: Z14967; MUID:98145437
 A;Accession: T03597
 A;Molecule type: mRNA; translated from GB/EMBL/DDBJ
 A;Residues: 1-304 <SUH>
 A;Cross-references: EMBL:Z93770; NID:g1929407; PIDN:CAB07805.1; PID:g1929408
 A;Experimental source: cultivar Xanthi.
 C;Genetics:
 C;Status: npp3
 C;Keywords: iron; phosphoprotein phosphatase; phosphoesterase core homology; phosphoprotein
 F;27-286/Domain: phosphoprotein phosphatase homology <PPP>
 F;55-123/Domain: phosphoesterase core homology <PBC>

Query Match 68.9%; Score 31; DB 2; Length 304;
 Best Local Similarity 66.7%; Pred. No. 39;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSVIQLSQ 10
 DB 22 GKLVQLSES 30
 |||||
 |||||

RESULT 3
 T24860
 hypothetical protein T12D8.4 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C;Accession: T24860
 similarity to the EMBL Data Library, October 1996
 A;Reference number: Z19944
 A;Accession: T24860
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-394 <WIL>
 A;Cross-references: EMBL:T281120; PIDN:CAB03344.1; GSPDB:GN00021; CESP:T12D8.4
 A;Experimental source: clone T12D8
 C;Genetics:
 A;Gene: CESP:T12D8.4
 A;Map position: 3
 A;Introns: 27/3; 58/1; 102/3; 159/3; 205/2; 289/3; 346/3

Query Match 68.9%; Score 31; DB 2; Length 394;
 Best Local Similarity 77.8%; Pred. No. 51;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
 DB 302 IGSIVQLSQ 310
 |||||
 |||||

RESULT 4
 D83498
 C4-dicarboxylate transport protein PA1183 [imported] - Pseudomonas aeruginosa (strain PA
 C;Species: Pseudomonas aeruginosa
 C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C;Accession: D83498
 R;Stover, C.K.; Egan, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A;Reference number: A82950; MUID:20437337

A;Accession: D83498
 A;Status: Preliminary
 A;Residues: 1-436 <STO>
 A;Cross-references: GB:AE004548; GB:AE004091; NID:g9947102; PIDN:AAG04572.1; GSPDB:GN
 A;Experimental source: strain PA01
 C;Genetics:
 C;Superfamily: C4-dicarboxylate carrier protein

Query Match 68.9%; Score 31; DB 2; Length 436;
 Best Local Similarity 66.7%; Pred. No. 57;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
 DB 214 VGSIVQLSQ 222
 |||||
 |||||

RESULT 5
 A49289
 Peroxisome proliferator-activated receptor - human
 C;Species: Homo sapiens (man)
 C;Date: 07-Jul-1999 #sequence_revision 07-Apr-1994 #text_change 20-Sep-1999
 C;Accession: A49289
 R;Sher, T.; Yi, H.F.; McBride, O.W.; Gonzalez, F.J.
 Biochemistry 32, 5598-5604, 1993
 A;Title: cDNA cloning, chromosomal mapping, and functional characterization of the hu
 A;Reference number: A49289; MUID:93277839
 A;Accession: A49289
 A;Molecule type: mRNA
 A;Cross-references: GB:L02932; NID:g307340; PIDN:AAA36468.1; PID:g307341
 A;Note: authors translated the codon AAC for residue 33 as Asp
 C;Genetics:
 A;Map position: 22
 A;Keywords: DNA binding; transcription regulation; zinc finger
 F;100-386/Domain: erba transforming protein homology <ERBA>

Query Match 68.9%; Score 31; DB 2; Length 468;
 Best Local Similarity 60.0%; Pred. No. 61;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
 DB 31 MGNIQEIQSQS 40
 |||||
 |||||

RESULT 6
 JC2085
 Peroxisome proliferator-activated receptor alpha chain - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 20-Sep-1999
 C;Accession: JC2085; S11659; S68199; S67963
 R;Gearring, K.L.; Crickmore, A.; Gustafsson, J.A.
 Biochem. Biophys. Res. Commun. 199, 255-263, 1994
 A;Title: Structure of the mouse peroxisome proliferator activated receptor alpha gene
 A;Reference number: JC2085; MUID:94108583
 A;Accession: JC2085
 A;Molecule type: DNA
 A;Residues: 1-468 <GEA>
 A;Cross-references: EMBL:X75287; EMBL:X75288; EMBL:X75289; NID:g467773; PIDN:CAA53042
 R;Isseemann, I.; Green, S.
 Nature 347, 645-650, 1990
 A;Title: The structure of the steroid hormone receptor superfamily by peroxi
 A;Reference number: S11659; MUID:91015382
 A;Accession: S11659
 A;Molecule type: mRNA
 A;Residues: 1-74,'R',76-468 <ISS>

A:Cross-references: GB:X57638; NID:g53764; PIDN:CAA0856.1; PID:g53765
R:Jones, P.S.; Savory, R.; Barratt, P.; Bell, A.R.; Gray, T.J.B.; Jenkins, N.A.; Gilbert
Eur. J. Biochem. 233, 219-226, 1995
A:Title: Chromosomal localisation, inducibility, tissue-specific expression and strain c
A:Reference number: S67963; MUID:96061953
A:Accession: S68199
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 413-468 <JON>
A:Cross-references: EMBL:X89577; NID:gl051294; PIDN:CAA61754.1; PID:gl051295
A:Experimental source: strain Swiss Webster; tissue liver
A:Accession: S67963
A:Molecule type: protein
A:Residues: 1-5 <JOW>
C:Comment: This protein mediates transcription of responsive genes which include those e
C:Genetics:
A:Gene: PPARalpha
A:Map position: 15
A:Introns: 70/1; 123/2; 170/1; 237/3; 387/1
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology
C:Keywords: DNA binding; receptor; transcription regulation; zinc finger
F:100-386/Domain: erba transforming protein homology <ERBA>
F:102-165/Region: DNA binding
F:139-156/Region: zinc finger CCCC motif
F:273-468/Domain: ligand binding #status predicted <LIB>

Query Match 68.9%; Score 31; DB 2; Length 468;
Best Local Similarity 60.0%; Pred. No. 61;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
||:|:||||
Db 31 MGNIQEIISQS 40

RESULT 7
A45288
peroxisome proliferator-activated receptor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 20-Sep-1999
C:Accession: A45288
R:Goettlicher, M.; Widmark, E.; Li, Q.; Gustafsson, J.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 4653-4657, 1992
A:Title: Fatty acids activate a chimera of the clofibrate acid-activated receptor and the
A:Reference number: A45288; MUID:92262498
A:Accession: A45288
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-468 <GOE>
C:Cross-references: GB:M88592; NID:g206317; PIDN:AAA41918.1; PID:g206318
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology
C:Keywords: DNA binding; transcription regulation; zinc finger
F:100-386/Domain: erba transforming protein homology <ERBA>

Query Match 68.9%; Score 31; DB 2; Length 468;
Best Local Similarity 60.0%; Pred. No. 61;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
||:|:||||
Db 31 MGNIQEIISQS 40

RESULT 8
I56603
peroxisome proliferator activated receptor alpha - human
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Sep-1999
C:Accession: I56603
R:Mukherjee, R.; Jow, L.; Noonan, D.; McDonnell, D.P.
J. Steroid Biochem. Mol. Biol. 51, 157-166, 1994

A:Title: Human and rat peroxisome proliferator activated receptors (PPARs) demonstrat
A:Reference number: I56603; MUID:95071923
A:Accession: I56603
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-468 <RES>
A:Cross-references: GB:S74349; NID:g765239; PIDN:AAB32649.1; PID:g765240
C:Genetics:
A:Gene: HPPARalpha
C:Superfamily: unassigned erba-related proteins; erba transforming protein homology
C:Keywords: zinc finger
F:100-386/Domain: erba transforming protein homology <ERBA>

Query Match 68.9%; Score 31; DB 2; Length 468;
Best Local Similarity 60.0%; Pred. No. 61;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
||:|:||||
Db 31 MGNIQEIISQS 40

RESULT 9
T46319
hypothetical protein DKFP434G0812.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46319
R:Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23035
A:Accession: T46319
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-617 <AAA>
A:Cross-references: EMBL:AL137619
A:Experimental source: adult testis; clone DKFP434G0812
C:Genetics:
A:Note: DKFP434G0812.1

Query Match 68.9%; Score 31; DB 2; Length 617;
Best Local Similarity 60.0%; Pred. No. 82;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVOLSQS 10
||:|:||||
Db 275 LGSLEQLSQA 284

RESULT 10
A38172
adenylate cyclase (EC 4.6.1.1) - Pasteurella multocida
C:Species: Pasteurella multocida
C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 24-Nov-1999
C:Accession: A38172
R:Mock, M.; Crasnier, M.; Duflot, E.; Dumay, V.; Danchin, A.
J. Bacteriol. 173, 6265-6269, 1991
A:Title: Structural and functional relationships between Pasteurella multocida and en
A:Reference number: A38172; MUID:92011391
A:Accession: A38172
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-838 <MOC>
A:Cross-references: GB:M68901; NID:gl50497; PIDN:AAA25532.1; PID:g551912
C:Superfamily: adenylate cyclase
C:Keywords: phosphorus-oxygen lyase

Query Match 68.9%; Score 31; DB 2; Length 838;
Best Local Similarity 60.0%; Pred. No. 11e+02;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSTVOLQS 10
 DB 104 MGSIASISQT 113

RESULT 11

C:Species: sucrose-phosphate synthase - fava bean (fragment)
 C:Species: Vicia faba (fava bean)
 C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 18-Jun-1999
 C:Accession: S53083
 R: Buchner, P.
 submitted to the EMBL Data Library, March 1995
 A:Description: Vicia faba
 A:Reference: M. J. M. S53083
 A:Accession: S53083
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-117 <BUC>
 A:Cross-references: EMBL:Z48640; NID:G732985; PIDN:CAA88587.1; PID:G732986
 C:Species: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
 F:1-117/Domain: sucrose/sucrose-phosphate synthase homology (fragment) <SPS>

Query Match 66.7%; Score 30; DB 2; Length 117;
 Best Local Similarity 50.0%; Pred. No. 25;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSTVOLQS 10
 DB 89 MGHIQMSKA 98

RESULT 12

C:Species: transcription elongation factor - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: C75331
 R: White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 A:Title: Complete genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896

A:Accession: C75331
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-161 <WHI>
 A:Cross-references: GB:AE002035; NID:G6459753; PIDN:AAF11523.1; PID:G645975
 C:Species: transcription elongation factor greb
 A:Map position: 1
 C:Superfamily: transcription elongation factor greb
 C:Keywords: transcription factor

Query Match 66.7%; Score 30; DB 2; Length 161;
 Best Local Similarity 75.0%; Pred. No. 34;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSTVOLQS 8
 DB 88 LGAIIVQLS 95

RESULT 13

S78130
 C:Species: ribosomal protein L10 - Reclinomonas americana (ATCC 50394) mitochondrion
 A:Variety: ATCC 50394

C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 05-Nov-1999
 C:Accession: S78130
 R: Lang, B.F.; Burger, G.; O'Kelly, C.J.; Cedergren, R.; Golding, G.B.; Lemieux, C.; S
 Nature 387, 493-497, 1997
 A:Title: An ancestral mitochondrial DNA resembling a eubacterial genome in miniature.

A:Reference number: S78127; MUID:97311393
 A:Accession: S78127
 A:Status: preliminary
 A:Molecule type: cDNA
 A:Residues: 1-193 <LAN>

A:Cross-references: EMBL:AF007261; NID:G2258325; PIDN:AAD11863.1; PID:G2258329
 A:Experimental source: ATCC 50394
 A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1997

A:Genetics: 110
 A:Genome: mitochondrion
 A:Start codon: GTG
 C:Keywords: mitochondrion; protein biosynthesis; ribosome

Query Match 66.7%; Score 30; DB 2; Length 193;
 Best Local Similarity 75.0%; Pred. No. 41;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 SIVOLQS 10
 DB 161 SVVOLQTS 168

RESULT 14

C83133
 C:Species: short-chain dehydrogenase PA4098 [imported] - Pseudomonas aeruginosa (strain
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: C83133
 R: Stoves, C.K.; Pham, X.O.; Brwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
 Smarby, C.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
 Lory, S.; Olson, M.V.
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 Nature 406, 959-964, 2000

A:Reference number: A82950; MUID:20437337
 A:Accession: C83133
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-241 <STO>
 A:Cross-references: GB:AE004826; NID:G9950296; PIDN:AAG07485.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4098
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 66.7%; Score 30; DB 2; Length 241;
 Best Local Similarity 66.7%; Pred. No. 52;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIVOLQS 10
 DB 151 GAIVQLTRS 159

RESULT 15

G81421
 C:Species: flagellar motor switch protein Cj0059c [imported] - Campylobacter jejuni (st
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: G81421
 R: Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference number: A81250; MUID:20150912
 A:Accession: G81421

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <PAR>
A:Cross-references: GB:AL139074; GB:AL111168; NID:g6967505; PIDN:CAB72546.1; PID:g696759
A:Experimental source: serotype O2, strain NCYC 11168
C:Genetics:
A:Gene: flvY; Cj0059c

Query Match 66.7%; Score 30; DB 2; Length 280;
Best Local Similarity 55.6%; Pred. No. 60;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 9
:||:|:|:
DB 227 IGSVVELNQ 235

RESULT 16
S32237
hypothetical protein 31 (rplJ 5' region) - Streptomyces griseus
C:Species: Streptomyces griseus
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 10-Mar-1994
C:Accession: S32237
R:Kuester, K.; Kuberski, S.; Piepersberg, W.; Distler, J.
submitted to the EMBL Data Library, March 1993
A:Description: Cloning and nucleotide sequence analysis of the nusG-rplK-rplA-rplJ-rplL
A:Reference number: S32234
A:Accession: S32237
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-305 <KUE>
A:Cross-references: EMBL:X72787

Query Match 66.7%; Score 30; DB 2; Length 305;
Best Local Similarity 50.0%; Pred. No. 66;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
|||:|:|:
DB 75 MGSVMSMKQS 84

RESULT 17
T48752
GRR1 related protein [imported] - Neurospora crassa (fragment)
N:Alternate names: protein 8D4.260
C:Species: Neurospora crassa
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: T48752
R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224541
A:Accession: T48752
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-364 <SCH>
A:Cross-references: EMBL:AL353819; GSPDB:GN00112; NCSP:8D4.260
A:Experimental source: cosmid contig 8D4; strain 74
C:Genetics:
A:Gene: NCSP:8D4.260
A:Map position: 2
A:Introns: 267/3

Query Match 66.7%; Score 30; DB 2; Length 364;
Best Local Similarity 60.0%; Pred. No. 79;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
:|:|:|:|:
DB 252 IASVLQAS 261

RESULT 18
E85615
probable transport ycad [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: E85615
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, E.; Potamouis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85615
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-382 <STO>
A:Cross-references: GB:AE005174; NID:gl2514063; PIDN:AAG55385.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ycad

Query Match 66.7%; Score 30; DB 2; Length 382;
Best Local Similarity 60.0%; Pred. No. 83;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
:||:|:|:
DB 277 LGSIALMSQA 286

RESULT 19
A64829
membrane protein ycad - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: A64829; S03788
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: A64829
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-382 <BLAT>
A:Cross-references: GB:AE000192; GB:U00096; NID:gl787125; PIDN:AAC73984.1; PID:gl7871
A:Experimental source: strain K-12, substrain MG1655
R:Blouin, P.T.; Cole, S.T.; Anderson, W.F.; Weiner, J.H.
Mol. Microbiol. 2, 785-795, 1988

A:Title: Nucleotide sequence of the dmsABC operon encoding the anaerobic dimethylsulph
A:Reference number: S03784; MUID:89096500
A:Accession: S03788
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-264, 'H' <BIL>
C:Genetics:
A:Gene: ycad

A:Map position: 20 min
C:Keywords: transmembrane protein
F:8-24/Domain: transmembrane #status predicted <TM1>
F:76-92/Domain: transmembrane #status predicted <TM2>
F:103-119/Domain: transmembrane #status predicted <TM3>
F:160-176/Domain: transmembrane #status predicted <TM4>
F:209-225/Domain: transmembrane #status predicted <TM5>
F:235-251/Domain: transmembrane #status predicted <TM6>
F:265-281/Domain: transmembrane #status predicted <TM7>
F:286-302/Domain: transmembrane #status predicted <TM8>
F:352-368/Domain: transmembrane #status predicted <TM9>

Query Match 66.7%; Score 30; DB 2; Length 382;

RESULT 24

T17339
 hypothetical protein DKFZp4340225.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T17339
 R:Duesternhoef, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, September 1999
 A:Reference number: Z18727
 A:Accession: T17339
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-662 <DUE>
 A:Cross-references: EMBL:AL117647
 A:Experimental source: adult testis; clone DKFZp4340225
 C:Genetics:
 A:Note: DKFZp4340225.1

Query Match 66.7%; Score 30; DB 2; Length 662;
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIVQLS 8
 II:IIII
 DB 590 GSVVQLS 596

RESULT 25
 S75636
 sensory transduction histidine kinase slr1969 - Synechocystis sp. (strain PCC 6803)
 N:Alternate names: protein slr1969
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C:Accession: S75636
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S.
 A:Reference number: S74322; MUID:97061201
 A:Accession: S75636
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-750 <KAN>
 A:Cross-references: EMBL:D90912; GB:AB001339; NID:g1653228; PIDN:RAA18197.1; PID:d101893
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: response regulator homology
 C:Keywords: phosphoprotein
 F:632-741/Domain: response regulator homology <RRH>
 F:680/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 66.7%; Score 30; DB 2; Length 750;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
 :II:IIII
 DB 437 INSIIQLSQ 445

RESULT 26
 B33926
 DNA-directed RNA polymerase (EC 2.7.7.6) chain A [validated] - Sulfolobus acidocaldarius
 C:Species: Sulfolobus acidocaldarius
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Aug-2000
 R:Puehler, G.; Leffers, H.; Gropp, F.; Palm, P.; Klenk, H.P.; Lottspeich, F.; Garrett, R.
 Proc. Natl. Acad. Sci. U.S.A. 86, 4569-4573, 1989
 A:Title: Archaeobacterial DNA-dependent RNA polymerases testify to the evolution of the e
 A:Reference number: A33926; MUID:89282812
 A:Accession: B33926

A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
 A:Molecule type: DNA
 A:Residues: 1-880 <PUE>
 R:Puehler, G.; Lottspeich, F.; Zillig, W.
 Nucleic Acids Res. 17, 4517-4534, 1989
 A:Title: Organization and nucleotide sequence of the genes encoding the large subunit
 submitted to the Protein Sequence Database, September 1999
 A:Reference number: S04714; MUID:89315197
 A:Accession: S04717
 A:Molecule type: DNA
 A:Residues: 1-311, 'N', 313-560, 'N', 562-610, 'M', 612-640, 'M', 642-880 <PUE>
 A:Cross-references: EMBL:X14818; NID:g46667; PIDN:CAA32925.1; PID:g46670
 C:Genetics:
 A:Gene: rpoA
 C:Function:
 A:Description: EC 2.7.7.6 [validated; MUID:89315197]
 C:Superfamily: Halobacterium DNA-directed RNA polymerase chain A
 C:Keywords: nucleotidyltransferase; transcription

Query Match 66.7%; Score 30; DB 1; Length 880;
 Best Local Similarity 77.8%; Pred. No. 2e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
 IIII:IIII
 DB 170 GSIVKLSPS 178

RESULT 27
 A84693
 hypothetical protein At2g29140 [Imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: A84693
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: A84693
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-964 <STO>
 A:Cross-references: GB:AE002093; NID:g3980417; PIDN:AA095220.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g29140
 A:Map position: 2

Query Match 66.7%; Score 30; DB 2; Length 964;
 Best Local Similarity 75.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQ 9
 I:IIII:II
 DB 846 GKIVQMSQ 853

RESULT 28
 F84693
 hypothetical protein At2g29200 [Imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F84693
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon,
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: F84693
 A:Status: preliminary

A:Molecule type: DNA
A:Accession: U956 <SPO>
A:Cross-references: GB:AE002093; NID:g3980412; PIDN:AAC95215.1; GSPDB:GN00139
C:Genetics: At2g29200
A:Gene: At2g29200
A:Map position: 2

Query Match 66.7%; Score 30; DB 2; Length 968;
Best Local Similarity 75.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSVVQLSQ 9
| | | | |
DB 850 GKIVQMSQ 857

RESULT 29
E84693
hypothetical protein At2g29190 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
A:Accession: E84693
R:Rounsley, S.D.; Shear, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Natter, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84693
A:Species: Arabidopsis thaliana
A:Molecule type: DNA
A:Cross-references: GB:AE002093; NID:g3980413; PIDN:AAC95216.1; GSPDB:GN00139
C:Genetics: At2g29190
A:Gene: At2g29190
A:Map position: 2

Query Match 66.7%; Score 30; DB 2; Length 972;
Best Local Similarity 75.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSVVQLSQ 9
| | | | |
DB 854 GKIVQMSQ 861

RESULT 30
T12195
sucrose-phosphate synthase (EC 2.4.1.14) - fava bean
C:Species: Vicia faba (fava bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T12195
R:Heim, U.; Weber, H.; Wobus, U.
Gene 178, 201-203, 1996
A:Title: Cloning and characterization of a full-length cDNA encoding sucrose phosphate
A:Reference number: Z17449; MUID:97080571
A:Accession: T12195
A:Species: Vicia faba
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1059 <HEI>
A:Cross-references: EMBL:Z56278; NID:g1022364; PIDN:CAA91217.1; PID:g1022365
A:Experimental source: cultivar Fribo, seed coat
C:Genetics:
A:Gene: SPS
A:Map position:
A:Description: catalyzes formation of sucrose-6-phosphate from UDPglucose and D-fructose
A:Pathway: sucrose biosynthesis
C:Superfamily: sucrose-phosphate synthase; sucrose/sucrose-phosphate synthase homology
C:Keywords: glycosyltransferase; hexosyltransferase; sucrose biosynthesis
F:169-654/domain: sucrose/sucrose-phosphate synthase homology <SSS>

Query Match 66.7%; Score 30; DB 2; Length 1059;
Best Local Similarity 50.0%; Pred. No. 2.4e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
| | | | |
DB 286 MGHIIQMSKA 295

RESULT 31
S53043
probable membrane protein YMR012W - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR270.16
C:Species: Saccharomyces cerevisiae
C:Accession: S53043 #sequence_revision 01-Sep-1995 #text_change 06-Feb-1998
R:Devlin, K.; Churcher, C.M.
submitted to the EMBL Data Library, March 1995
A:Reference number: S53028
A:Accession: S53043
A:Molecule type: mRNA
A:Cross-references: EMBL:Z48613; NID:g728645; PID:g728661; MIPS:YMR012W
A:Experimental source: strain AB972
C:Genetics:
A:Gene: SGD:CLD1
A:Cross-references: SGD:S0004614; MIPS:YMR012W
C:Map position: 13R
A:Species: Saccharomyces cerevisiae
F:855-851/domain: transmembrane #status predicted <TMM>

Query Match 66.7%; Score 30; DB 2; Length 1277;
Best Local Similarity 55.6%; Pred. No. 2.9e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
| | | | |
DB 710 LGKIIELSQ 718

RESULT 32
T17285
hypothetical protein DKFZp434N0535.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17285
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Accession: T17285
A:Cross-references: Z18723
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1341 <POU>
A:Cross-references: EMBL:AL117518
A:Experimental source: adult testis; clone DKFZp434N0535
A:Note: DKFZp434N0535.1

Query Match 66.7%; Score 30; DB 2; Length 1341;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSVVQLS 8
| | | | |
DB 1269 GSVVQLS 1275

RESULT 33
T23620
hypothetical protein K12D12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
C:Accession: T23620; T28109
R:Colles, L.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z19772
A:Accession: T23620
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1520 <WIL>
A:Cross-references: EMBL:Z49069; PIDN:CAA88867.1; GSPDB:GN000020; CESP:K12D12.1
A:Experimental source: clone K12D12
R:Swinburne, J.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20470
A:Accession: T28109
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1520 <W12>
A:Cross-references: EMBL:Z70213; PIDN:CAA94177.1; GSPDB:GN000020; CESP:K12D12.1
A:Experimental source: clone ZK930
C:Genetics:
A:Gene: CESP:K12D12.1
A:Map position: 2
A:Introns: 34/1; 146/2; 390/3; 471/3; 611/2; 1351/3; 1486/2
C:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hyd

Query Match 66.7%; Score 30; DB 2; Length 1520;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
II:II I:I
Db 800 MGTIVNLAQ 808

RESULT 34
T09144
probable guanine nucleotide exchange factor RhoGEF2 - fruit fly (Drosophila melanogaster)
N:Alternate names: Shar pei/DRHOGEF2
C:Species: Drosophila melanogaster
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Nov-2000
C:Accession: T09144; T09223
R:Haecker, U.; Perrimon, N.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z16586
A:Accession: T09144
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2559 <HAE>
A:Cross-references: EMBL:AF031930; NID:g2687355; PID:g2687356
R:Barrett, K.; Leptin, M.; Settleman, J.
Cell 91, 905-915, 1997
A:Title: The Rho GTPase and a putative RhoGEF mediate a signaling pathway for the cell
A:Reference number: Z16618; MUID:98088790
A:Accession: T09223
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-676 'L', 678-837, 'L', 839-889, 'S', 891-1243, 'D', 1245-1358, 'E', 1360-1368, 'R', 1
A:Cross-references: EMBL:AF032870; NID:g2760367; PIDN:AAC38820.1; PID:g2760368
C:Genetics:
A:Gene: rhoGEF2
A:Cross-references: FlyBase:FBgn023172
A:Map position: 2; 53F1-2
A:Note: orchestrates cell shape changes during gastrulation
C:Function:
A:Description: mediates actin rearrangements required for cell shape changes during gastrulation
C:Superfamily: protein kinase C zinc-binding repeat homology
C:Keywords: signal transduction; embryo; GTP exchange
F:1151-1200/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 66.7%; Score 30; DB 2; Length 2559;
Best Local Similarity 70.0%; Pred. No. 5.9e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
II III:II
Db 196 MGGVQLNQS 205

RESULT 35
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C:Species: Amycolatopsis mediterranei
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C:Accession: T17464
R:Schüpp, T.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z18802
A:Accession: T17464
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-5069 <SCH>
A:Cross-references: EMBL:AJ223012; NID:el227119; PID:el227121; PIDN:CAA11036.1
A:Experimental source: strain LBG A3136
C:Superfamily: acyl carrier protein homology
C:Keywords: carrier protein
F:1631-1702/Domain: acyl carrier protein homology <ACP1>
F:2238-3309/Domain: acyl carrier protein homology <ACP2>
F:4939-5010/Domain: acyl carrier protein homology <ACP3>

Query Match 66.7%; Score 30; DB 2; Length 5069;
Best Local Similarity 50.0%; Pred. No. 1.2e+03;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
II:II:II
Db 4219 LGAVVELPQS 4228

RESULT 36
S12393
hypothetical protein (gacA 5' region) - Staphylococcus aureus plasmid pSK1
C:Species: Staphylococcus aureus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S12393
R:Rouch, D.A.; Cram, D.S.; DiBerardino, D.; Littlejohn, T.G.; Skurray, R.A.
Mol. Microbiol. 4, 2051-2062, 1990
A:Title: Efflux-mediated antiseptic resistance gene qacA from Staphylococcus aureus:
A:Reference number: S12393; MUID:91211614
A:Accession: S12393
A:Molecule type: DNA
A:Residues: 1-188 <ROU>
A:Cross-references: EMBL:X56628; NID:g773395; PIDN:CAA39962.1; PID:g46660
C:Genetics:
A:Genome: plasmid

Query Match 64.4%; Score 29; DB 2; Length 188;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
I II:II
Db 26 GEIVKLSES 34

RESULT 37
G86748
hypothetical protein ykcB [imported] - Lactococcus lactis subsp. lactis (strain IL140
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001
C:Accession: G86748

R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malmgren, K.; Weissenbach, J.; Ehrlich
 Genome Res. In press, 2001
 A:Reference: complete genome sequence of the lactic acid bacterium.
 A:Accession: G86748

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-225 <STO>
 A:Cross-references: GB:AA005176; NID:gl272935; PIDN:AAK05089.1; GSPDB:GN00146
 A:Experimental source: strain IL403
 C:Genetics:
 A:Gene: ykcb

Query Match 64.4%; Score 29; DB 2; Length 225;
 Best Local Similarity 75.0%; Pred. No. 81;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 MGSIVQLSQ 9
 Db 198 GSIIQKSO 205

RESULT 38
 F84145
 transposase (23) BH3966 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: F84145
 R;Takami, T.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20263314
 A:Accession: F84145
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-234 <STO>
 A:Cross-references: GB:AP001520; GB:BA000004; NID:gl0176401; PIDN:BAB07685.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3966
 C:Superfamily: DNA replication protein dnaC

Query Match 64.4%; Score 29; DB 2; Length 254;
 Best Local Similarity 60.0%; Pred. No. 91;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGSIVQLSQS 10
 Db 137 MGEIVQLKLS 146

RESULT 39
 H84149
 transposase (27) BH4000 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: H84149
 R;Takami, T.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20263314
 A:Accession: H84149
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2561 <STO>
 A:Cross-references: GB:AP001520; GB:BA000004; NID:gl0176401; PIDN:BAB07719.1; GSPDB:GN00
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH4000
 C:Superfamily: DNA replication protein dnaC

Query Match 64.4%; Score 29; DB 2; Length 261;
 Best Local Similarity 60.0%; Pred. No. 94;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGSIVQLSQS 10
 Db 144 MGEIVQLKLS 153

RESULT 40
 T29574
 hypothetical protein F30B5.7 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29574
 R;Haw, H.
 submitted to the EMBL Data Library, December 1995
 A:Description: The sequence of C. elegans cosmid F30B5.
 A:Reference number: 220644
 A:Accession: T29574
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-74
 A:Cross-references: EMBL:U42437; PIDN:AAA83496.1; CESP:F30B5.7
 C:Genetics:
 A:Gene: CESP:F30B5.7
 A:Introns: 33/3; 79/2; 156/3

Query Match 64.4%; Score 29; DB 2; Length 274;
 Best Local Similarity 60.0%; Pred. No. 99;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGSIVQLSQS 10
 Db 177 MGSILQNSSET 186

RESULT 41
 H71938
 flagellar motor switch protein - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C:Accession: H71938
 R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.
 ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
 A:Reference number: A71800; MUID:99120557
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-285 <ARN>
 A:Cross-references: GB:AE001473; GB:AE001439; NID:94154910; PIDN:AA05971.1; PID:9415
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp0394

Query Match 64.4%; Score 29; DB 2; Length 285;
 Best Local Similarity 55.6%; Pred. No. 1e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGSIVQLSQ 9
 Db 231 IGSVVQLDQ 239

RESULT 42
 F04648
 flag protein - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: F64648
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliak, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C. Nature 388, 539-547, 1997
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: F64648
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-287 <TOM>
A:Cross-references: GB:AE000611; GB:AE000511; NID:g23114173; PIDN:AAD08074.1; PID:g2311417

Query Match 64.4%; Score 29; DB 2; Length 287;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 MGSIVQLSQ 9
:||:|:|:
Db 233 IGSVVELDQ 241

RESULT 43
A40573
clathrin heavy chain - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 13-Aug-1999
C:Accession: A40573
R:Dodge, G.R.; Kovalsky, I.; McBride, O.W.; Yi, H.F.; Chu, M.; Saitta, B.; Stokes, D.G. Genomics 11, 174-178, 1991
A:Title: Human clathrin heavy chain (CLTC): partial molecular cloning, expression, and mapping to chromosome 16p11.
A:Reference number: A40573; MUID:92112210
A:Accession: A40573
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-305 <DOD>
A:Cross-references: GB:X55878; GB:S75467; NID:g29982; PIDN:CAA39363.1; PID:g29983
C:Superfamily: clathrin heavy chain

Query Match 64.4%; Score 29; DB 2; Length 305;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 MGSIVQLSQ 9
:||||:|:
Db 156 LGSIVNFSQ 164

RESULT 44
H70790
hypothetical protein Rv3679 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70790
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: H70790
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-340 <COL>
A:Cross-references: GB:AL022121; GB:AL123456; NID:g3261559; PIDN:CAA18001.1; PID:el26453
A:Experimental source: strain H37RV
C:Genetics:

A:Gene: Rv3679

Query Match 64.4%; Score 29; DB 2; Length 340;
Best Local Similarity 50.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 MGSIVQLSQS 10
:||:|:|:
Db 324 LGSLYELSES 333

RESULT 45
A85877
probable prophage DNA injection protein Z3614 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: A85877
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May, J.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apoda, N. Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85877
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <STO>
A:Cross-references: GB:AE005174; NID:g12516716; PIDN:AAG57477.1; GSPDB:GN00145; UWGP:157
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3614

Query Match 64.4%; Score 29; DB 2; Length 378;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 MGSIVQLSQS 10
:||:|:|:
Db 59 LGSVAQTSQA 68

Search completed: June 28, 2001, 11:57:06
Job time: 247 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2001, 11:57:29 ; Search time 22.61 seconds
(without alignments)
15.151 Million cell updates/sec

Title: US-09-439-313-573

Perfect score: 45

Sequence: 1 MGSIVQLSQS 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : SwissProt_39:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	31	68.9	304	1 PPI3_TOBAC	O04858 nicotiana t
2	31	68.9	428	1 DCTA_SALTY	P50334 salmonella
3	31	68.9	468	1 PPAR_HUMAN	O07869 homo sapien
4	31	68.9	468	1 PPAR_MOUSE	P23204 mus musculus
5	31	68.9	468	1 PPAR_RAT	P37230 rattus norv
6	31	68.9	490	1 C7DB_LOTJA	O22307 lotus japon
7	31	68.9	838	1 CYAA_PASMO	O05766 pasteurilla
8	31	68.9	1508	1 AT5A_MOUSE	O54927 mus musculus
9	30	66.7	305	1 Y031_STRGR	P36261 streptomyce
10	30	66.7	316	1 PPI7_ARATH	O82733 arabidopsis
11	30	66.7	382	1 YCAD_ECOLI	P21503 escherichia
12	30	66.7	657	1 DCTS_RHOCA	P37739 rhodobacter
13	30	66.7	880	1 RPA1_SULAC	P11512 sulfolobus
14	30	66.7	1059	1 SPS_VICFA	O43876 vicia faba
15	30	66.7	1277	1 IF3X_YEAST	O03690 saccharomyc
16	30	66.7	1520	1 TOP2_CAEEL	Q23670 caenorhabdi
17	29	64.4	188	1 YP23_STAAL	Q23670 caenorhabdi
18	29	64.4	340	1 Y0H9_MYCTU	P23217 staphylococ
19	29	64.4	384	1 NAPP_YEAST	O69647 mycobacteri
20	29	64.4	459	1 DHE4_EWENI	P53164 saccharomyc
21	29	64.4	463	1 SELA_ECOLI	P18819 emericalla
22	29	64.4	503	1 VGLY_P1ARV	P23328 escherichia
23	29	64.4	511	1 PUR1_DROME	P03540 pichinde ar
24	29	64.4	540	1 KNL_CAEEL	P46822 caenorhabdi
25	29	64.4	622	1 SKN7_YEAST	P38889 saccharomyc
26	29	64.4	1171	1 DIA3_MOUSE	Q92207 mus musculu
27	29	64.4	1640	1 CLH2_HUMAN	P53675 homo sapien
28	29	64.4	1675	1 CLH1_HUMAN	Q00610 homo sapien
29	29	64.4	1675	1 CLH_BOVIN	P49591 bos taurus
30	29	64.4	1675	1 CLH_RAT	P11442 rattus norv
31	29	64.4	1681	1 CLH_CAEEL	P34574 caenorhabdi
32	29	64.4	2787	1 TEL1_YEAST	P38110 saccharomyc
33	28	62.2	337	1 G3P_MYCGE	P47543 mycoplasma

ALIGNMENTS

RESULT 1

PP13_TOBAC	PP13_TOBAC	STANDARD;	PRT;	304 AA.
ID	PP13_TOBAC	STANDARD;	PRT;	304 AA.
AC	O04858;			
DT	01-OCT-2000 (Rel. 40, Created)			
DT	01-OCT-2000 (Rel. 40, Last sequence update)			
DT	01-OCT-2000 (Rel. 40, Last annotation update)			
DE	SERINE/THREONINE PROTEIN PHOSPHATASE PPI ISOZYME 3 (EC 3.1.3.16).			
GN	NPP3.			
OS	Nicotiana tabacum (Common tobacco).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;			
OC	Solanales; Solanaceae; Nicotiana.			
OX	NCBI_TaxID=4097;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. XANTHI.			
RX	MEDLINE=98145437; PubMed=9484443;			
RA	Sun M., Cho H., Kim Y., Liu J., Lee H.;			
RT	*Multiple genes encoding serine/threonine protein phosphatases and			
RT	their differential expression in Nicotiana tabacum.*;			
RL	Plant Mol. Biol. 36:315-322(1998).			
CC	-I- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +			
CC	ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC).			
CC	-I- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES. PP-1			
CC	SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z93770; CAB07805.1; -			

HSP; P08129; IEFM.
DR Model; 16119; NUCts;1262;16119.
DR InterPro; IPRO00934; -
DR Pfam; PF00149; Sphosphatase; 1.
DR PRINTS; PRO0114; STPHPTWASE.
DR PROSITE; PS00125; SER_THR_PHOSPHATASE; 1.
KW Hydroxylase; Iron; Manganese; Multigene family.
FT METAL 61
FT METAL 61
FT METAL 89
FT METAL 89
FT METAL 121
FT ACT_SITE 121
FT ACT_SITE 122
FT ACT_SITE 170
FT METAL 170
FT METAL 245
FT METAL 245
SQ SEQUENCE 304 AA; 34441 MW; DBA3DSD594857A452 CRC64;

Query Match: 68.9%; Score 31; DB 1; Length 304;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 1; Indels 0; Gaps 0;
OY 2 GSIVQLSQS 10
DB 22 GLRVQLSES 30

RESULT 2
ID DCTA_SALT STANDARD; PRT: 428 AA.
AC F50334.1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE C4-DICARBOXYLATE TRANSPORT PROTEIN.
GN DCTA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Bacteroidetes; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_1-taxid=602;
RN [1]_taxid=602;
RP SEQUENCE FROM N.A.
STRAINE-LT2;
RC MEDLINE-97113529; PubMed-8955389;
CC Baker K.E., Ditullio K.P., Neuhaud J., Kelln R.A.;
CC Utilization of orotate as a pyrimidine source by Salmonella
CC typhimurium and Escherichia coli requires the dicarboxylate transport
CC protein.
CC J. Bacteriol. 178:7099-7105(1996).
CC
CC !- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF DICARBOXYLATES SUCH AS
CC SUCCINATE, FUMARATE, AND MALATE FROM THE PERIPLASM ACROSS THE
CC INNER MEMBRANE (BY SIMILARITY).
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC !- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTEF FAMILY
CC (SDF).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announcement/
CC 19960101)-----
CC
CC EMBL; X91397; CAA62742.1; -
CC StyGene; SG10618; dcta.
DR InterPro; IPRO01991; -
DR Pfam; PF00375; SDF; 1.
DR PROSITE; PS00713; NA_DICARBOXYL_SYM_1; 1.
DR PROSITE; PS00713; NA_DICARBOXYL_SYM_2; 1.
KW Transporter; Transmembrane; Inner membrane; Symport.
FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 21 39 1 (POTENTIAL).
FT DOMAIN 40 58 PERIPLASMIC (POTENTIAL).

TRANSSEM 59 77 2 (POTENTIAL).
FT TRANSSEM 78 99 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 100 108 3 (POTENTIAL).
FT TRANSSEM 109 139 PERIPLASMIC (POTENTIAL).
FT TRANSSEM 140 158 4 (POTENTIAL).
FT TRANSSEM 159 161 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 162 180 5 (POTENTIAL).
FT TRANSSEM 181 199 PERIPLASMIC (POTENTIAL).
FT TRANSSEM 200 218 6 (POTENTIAL).
FT TRANSSEM 219 231 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 232 250 7 (POTENTIAL).
FT TRANSSEM 251 271 PERIPLASMIC (POTENTIAL).
FT TRANSSEM 272 290 8 (POTENTIAL).
FT TRANSSEM 291 310 9 (POTENTIAL).
FT TRANSSEM 311 329 10 (POTENTIAL).
FT TRANSSEM 330 349 11 (POTENTIAL).
FT TRANSSEM 360 368 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 369 388 11 (POTENTIAL).
FT TRANSSEM 389 395 PERIPLASMIC (POTENTIAL).
FT TRANSSEM 396 414 12 (POTENTIAL).
FT TRANSSEM 415 428 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 428 AA; 45449 MW; 86CE44C25539B48C CRC64;

Query Match: 68.9%; Score 31; DB 1; Length 428;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 6; Conservative 1; Indels 0; Gaps 0;
OY 1 MGSIQVLSQ 9
DB 213 VGSLVQLGQ 221

RESULT 3
PPAR_HUMAN STANDARD; PRT: 468 AA.
ID PPARG.HUMAN Q16241; Q92486;
AC Q07869; Q92489; Q16241; Q92486;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR ALPHA (PPAR-ALPHA).
GN PPARG OR NR1C1 OR PPAR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX [1]_taxid=9606;
RN [1]_taxid=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE-93277839; PubMed-7684926;
RA Sher T., Yi H.P., McBride O.W., Gonzales F.J.;
RT "cDNA cloning, chromosomal mapping, and functional characterization
RT of the human peroxisome proliferator activated receptor.";
RB Biochemistry 32:5598-5604(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC SEQUENCE FROM N.A.
RX Roberts R.A., James N.H., Woodruff N.J., Macdonald N., Tugwood J.D.;
RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC SEQUENCE FROM N.A.
RX Mukherjee R., Jow L., Noonan D., McDonnell D.P.;
RA "Human and rat peroxisome proliferator activated receptors (PPARs)
RT demonstrate similar tissue distribution but different responsiveness
RT to PPAR activators.";
RB J. Steroid Biochem. Mol. Biol. 51:157-166(1994).
RN [4]
RP SEQUENCE OF 124-468 FROM N.A.
RC Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RA Submitter: RECEPTOR THAT BIND PEROXISOME PROLIFERATORS SUCH AS

CC HYPOLIPIDEMIC DRUGS AND FATTY ACIDS. ONCE ACTIVATED BY A LIGAND,
CC THE RECEPTOR BINDS TO A PROMOTER ELEMENT IN THE GENE FOR ACYL-COA
CC OXIDASE AND ACTIVATES ITS TRANSCRIPTION. IT THEREFORE CONTROLS THE
CC PEROXISOMAL BETA-OXIDATION PATHWAY OF FATTY ACIDS.
CC -1- SUBUNIT: HETERODIMER WITH THE RETINOID X RECEPTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L02932; AAA36468.1; -
CC EMBL; Y07619; CAA68898.1; -
CC EMBL; S74349; AAB32649.1; -
CC EMBL; AL078611; CAB44427.1; -
CC PIR; A49289; A49289.
CC HSP; P03372; IHCO.
CC MIM; 170998; -
CC InterPro; IPR000536; -
CC InterPro; IPR001628; -
CC InterPro; IPR001723; -
CC InterPro; IPR003074; -
CC InterPro; IPR003076; -
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00047; STRIDFINGER.
CC PRINTS; PR00398; STRDHORMONER.
CC PRINTS; PR01288; PROXISOMPAR.
CC PRINTS; PR01289; PROXISOMPAR.
CC PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
CC Receptor; Transcription regulation; Activator; DNA-binding;
CC Nuclear protein; Zinc-finger; Multigene family.
CC DNA_BIND 102 166 NUCLEAR RECEPTOR-TYPE.
CC ZN_FING 102 122 C4-TYPE.
CC ZN_FING 139 161 C4-TYPE.
CC DOMAIN 281 468 LIGAND-BINDING (POTENTIAL).
CC CONFLICT 71 71 T -> M (IN REF. 2).
CC CONFLICT 123 123 K -> M (IN REF. 2).
CC CONFLICT 268 268 A -> V (IN REF. 1).
CC CONFLICT 296 296 G -> A (IN REF. 1).
CC CONFLICT 444 444 V -> A (IN REF. 2).
CC SEQUENCE 468 AA; 52225 MW; 850846FD51ADA883 CRC64;
Query Match 68.9%; Score 31; DB 1; Length 468;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MGSIVQLSQS 10
Db 31 MGNIQEISQS 40
||:|:||||
RESULT 4
PPAR_MOUSE
ID PPAR_MOUSE STANDARD; PRT; 468 AA.
AC P23204;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PEROXISOME PROLIFERATOR ACTIVATED RECEPTOR ALPHA (PPAR-ALPHA).
GN PPARA OR NR1C1 OR PPAR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE-91015382; PubMed-2129546;
RA Issemann I., Green S.;
RT "Activation of a member of the steroid hormone receptor superfamily
RL by peroxisome proliferators.";
RN Nature 347:645-650(1990).
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE-94168583; PubMed-8123021;
RA Gearing K.L., Crickmore A., Gustafsson J.-A.;
RT "Structure of the mouse peroxisome proliferator activated receptor
RL alpha gene.";
RN Biochem. Biophys. Res. Commun. 199:255-263(1994).
RP SEQUENCE OF 413-468 FROM N.A.
RC STRAIN=SWISS WEBSTER; TISSUE=Liver;
RX MEDLINE-96061953; PubMed-7588749;
RA Jones P.S., Savory R., Barratt P., Bell A.R., Gray T.J.B.,
RJenkins N.A., Gilbert D.J., Copeland N.G., Bell D.R.;
RT "Chromosomal localisation, inducibility, tissue-specific expression
RL and strain differences in three murine
RP peroxisome-proliferator-activated-receptor genes.";
RJ Eur. J. Biochem. 233:219-226(1995).
CC -1- FUNCTION: RECEPTOR THAT BIND PEROXISOME PROLIFERATORS SUCH AS
CC HYPOLIPIDEMIC DRUGS AND FATTY ACIDS. ONCE ACTIVATED BY A LIGAND,
CC THE RECEPTOR BINDS TO A PROMOTER ELEMENT IN THE GENE FOR ACYL-COA
CC OXIDASE AND ACTIVATES ITS TRANSCRIPTION. IT THEREFORE CONTROLS THE
CC PEROXISOMAL BETA-OXIDATION PATHWAY OF FATTY ACIDS.
CC -1- SUBUNIT: HETERODIMER WITH THE RETINOID X RECEPTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER, KIDNEY AND
CC HEART. VERY WEAKLY EXPRESSED IN BRAIN AND TESTIS.
CC -1- DEVELOPMENTAL STAGE: IT APPEARS FIRST AT DAY 13.5 POSTCONCEPTION,
CC AND INCREASES UNTIL BIRTH.
CC -1- DISEASE: PEROXISOME PROLIFERATORS ARE A DIVERSE GROUP OF
CC CHEMICALS THAT INCLUDE HYPOLIPIDAEMIC DRUGS, HERBICIDES AND
CC INDUSTRIAL PLASTICISERS. ADMINISTRATION OF THESE CHEMICALS TO
CC RODENTS RESULTS IN THE DYNAMIC PROLIFERATION OF HEPATIC
CC PEROXISOMES AS WELL AS LIVER HYPERPLASIA.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X57638; CAA40856.1; -
CC EMBL; X75289; CAA53042.1; -
CC EMBL; X75290; CAA53042.1; JOINED.
CC EMBL; X75291; CAA53042.1; JOINED.
CC EMBL; X75292; CAA53042.1; JOINED.
CC EMBL; X75293; CAA53042.1; JOINED.
CC EMBL; X75294; CAA53042.1; JOINED.
CC EMBL; X89577; CAA61754.1; -
CC PIR; S11659; S11659.
CC HSP; P03372; IHCO.
CC TRANSFAC; T00694; -
CC MGD; MGI:104740; Ppara.
CC InterPro; IPR000536; -
CC InterPro; IPR001628; -
CC InterPro; IPR001723; -
CC InterPro; IPR003074; -
CC InterPro; IPR003076; -
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00047; STRIDFINGER.
CC PRINTS; PR00398; STRDHORMONER.
CC PRINTS; PR01288; PROXISOMPAR.

```

DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1
KW RECEPTOR; Transcription regulation; Activator; DNA-binding;
KW Nuclear protein; Zinc-finger; Multigene family.
KW DNA_BIND 102 166 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 102 132 C4-TYPE.
FT ZN_FING 139 161 C4-TYPE.
FT DOMAIN 281 468 LIGAND-BINDING (POTENTIAL).
SQ SEQUENCE 468 AA; 52377 MW; 2A89E7D71SC8DBA9 CRC64;

Query Match      68.9%; Score 31; DB 1; Length 468;
Best Local Similarity 60.0%; Pred. NO. 19;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVLSQS 10
DB 31 MGNIQISQS 40

RESULT 6
C7DB_LOTJA STANDARD; PRT; 490 AA.
ID C7DB_LOTJA
AC Q22307; 1998 (Rel. 37, Created)
DT 15-DEC-1998
DE LOTUS japonicus (date update)
FL 15-DEC-1998 (Rel. 37, Last annotation update)
DE CYTOCHROME P450 71D11 (EC 1.14.-.-) (FRAGMENT).
GN CYP71D11
OS Lotus japonicus.
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;
OC Fabales; Papilionaceae; Papilionoideae; Lotus.
KB KEGG; TraX; P43005;
RN 111_TraX; P43005;
RP SEQUENCE FROM N.A.
RC STRAIN=CV. GIFU; TISSUE=Root nodules;
RX MEDLINE=97422886; PubMed=9276951;
RA Szczyglowski K., Hamburger D., Kapranov P., de Bruijn F.J.;
RL Construction of a Lotus japonicus late nodulin expressed sequence
RT Plant Physiol 114:1335-1346(1997)
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use, provided the source is acknowledged. The information contained in this
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF000403; AAB69644.1;
CC InterPro; IPR001128; 1
CC Pfam; PF00060; P450CYTOCHROME_P450; 1.
CC Oxidoreductase; Monooxygenase; Heme.
KW NON_TER 1
FT BINDING 427 427 HEME (BY SIMILARITY).
SQ SEQUENCE 490 AA; 56023 MW; B6C83F2129D58247 CRC64;

Query Match      68.9%; Score 31; DB 1; Length 490;
Best Local Similarity 60.0%; Pred. NO. 20;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSVIVLSQS 10
DB 152 GSVVNLSQL 160

RESULT 7
CYAA_PASMU STANDARD; PRT; 838 AA.
ID CYAA_PASMU

```

AC Q05766;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE ADENYLATE CYCLASE (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE) (ADENYLYL
 DE CYCLASE).
 DE CYA.
 GN Pasteurella multocida.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CNP1 / NTCC 10322;
 RX MEDLINE=92011391; PubMed=1917858;
 RA Mock M., Crasner M., Duflot E., Dumay V., Danchin A.;
 RT "Structural and functional relationships between Pasteurella
 RT multocida and enterobacterial adenylate cyclases.";
 RL J. Bacteriol. 173:6265-6269(1991).
 RN [2]
 RP REVIEW.
 RX MEDLINE=93119764; PubMed=8418825;
 RA Danchin A.;
 RT "Phylogeny of adenylate cyclases.";
 RL Adv. Second Messenger Phosphoprotein Res. 27:109-162(1993).
 CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO ADENYLATE CYCLASE CLASS-1 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M68901; AAA25532.1; -;
 DR PIR: A38172; A38172.
 DR InterPro: IPR000274; -;
 DR Pfam: PF01295; Adenylate_cyclase_1.
 DR PROSITE: PS01092; ADENYLATE_CYCLASE_1_1; 1.
 DR PROSITE: PS01093; ADENYLATE_CYCLASE_1_2; 1.
 KW Lyase; CAMP synthesis.
 FT DOMAIN 1 541 CATALYTIC (POTENTIAL).
 FT DOMAIN 547 838 REGULATORY (POTENTIAL).
 FT SEQUENCE 838 AA; 96816 MW; 2DDFB43AC08C29C6 CRC64;
 SQ
 Query Match 68.9%; Score 31; DB 1; Length 838;
 Best Local Similarity 60.0%; Pred. No. 37;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 OY 1 MGSIVQLSQS 10
 IIII :II:
 DB 104 MGSIASISQT 113
 RESULT 8
 ID AT5A MOUSE STANDARD; PRT; 1508 AA.
 AC O54827;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE POTENTIAL PHOSPHOLIPID-TRANSPORTING ATPASE VA (EC 3.6.1.-).
 GN ATPC5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISUE=Teratocarcinoma;
 RX MEDLINE=20473714; PubMed=11015572;
 RA Halleck M.S., Lawler J.F. Jr., Blackshaw S., Gao L., Nagarajan P.,
 RA Hacker C., Pyle S., Newman J.T., Nakanishi Y., Ando H., Weinstein D.,
 RA Williamson P., Schlegel R.A.;
 RT "Differential expression of putative transbilayer amphipath
 RT transporters.";
 RL Physiol. Genomics 1:139-150(1999).
 RN [2]
 RP SEQUENCE OF 16-435 FROM N.A.
 RX MEDLINE=98217376; PubMed=9548971;
 RA Halleck M.S., Pradhan D., Blackman C.F., Berkes C., Williamson P.L.,
 RA Schlegel R.A.;
 RT "Multiple members of a third subfamily of p-type ATPases identified by
 RT genomic sequences and ESTs.";
 RL Genome Res. 8:354-361(1998).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: FOUND IN TESTIS. ALSO DETECTED IN FETAL
 CC TISSUES.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
 CC ATPASES). SUBFAMILY IV.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: AF011337; AAC02902.1; -;
 DR EMBL: AF156549; AAF09447.1; -;
 DR MGD; MGI:1330809; Atpc5.
 DR PROSITE: PS00154; ATPASE_E1_E2; 1.
 KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
 KW Multigene family.
 FT DOMAIN 1 79 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 80 101 POTENTIAL.
 FT DOMAIN 102 107 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 108 129 POTENTIAL.
 FT DOMAIN 130 313 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 314 335 POTENTIAL.
 FT DOMAIN 336 366 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 367 388 POTENTIAL.
 FT DOMAIN 389 1101 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1102 1122 POTENTIAL.
 FT DOMAIN 1123 1134 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1135 1154 POTENTIAL.
 FT DOMAIN 1155 1184 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1185 1206 POTENTIAL.
 FT DOMAIN 1207 1213 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1214 1236 POTENTIAL.
 FT DOMAIN 1237 1242 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 1243 1263 POTENTIAL.
 FT DOMAIN 1264 1281 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1282 1306 POTENTIAL.
 FT DOMAIN 1307 1508 CYTOPLASMIC (POTENTIAL).
 FT MOD_RES 431 431 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 1045 1045 MAGNESIUM (BY SIMILARITY).
 FT METAL 1049 1049 MAGNESIUM (BY SIMILARITY).
 FT DOMAIN 17 23 POLY-ARG.
 FT DOMAIN 471 474 POLY-GLU.
 FT CONFLICT 16 22 WRRPRR -> KLAARK (IN REF. 2).
 FT CONFLICT 435 435 T -> L (IN REF. 2).
 FT SEQUENCE 1508 AA; 168699 MW; DC9A0D99AF7EEB9E CRC64;
 SQ
 Query Match 68.9%; Score 31; DB 1; Length 1508;
 Best Local Similarity 60.0%; Pred. No. 71;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 OY 1 MGSIVQLSQS 10


```

RN  [3]
RC  SEQUENCE OF 1-265 FROM N.A.
RX  STRAIN-K12 / C600;
RA  MEDLINE-890956500; PubMed-3062312;
RT  Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H.;
RL  "Nucleotide sequence of the dmsABC operon encoding the anaerobic
    dimethylsulphoxide reductase of Escherichia coli.";
CC  Mol. Microbiol. 2:785-795(1988).
CC  -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE000192; AAC73984.1; -
DR  EMBL; D90727; BAA35630.1; -
DR  EMBL; D90728; BAA35633.1; -
DR  EMBL; J03412; AAA83847.1; -
DR  PIR; S03788; S03788.
DR  EcoGene; EG11242; ycad.
DR  InterPro; IPR001066; -.
DR  Pfam; PF00083; sugar_tr; 1.
KW  Hypothetical protein; Transmembrane.
FT  TRANSMEM 14 34 POTENTIAL.
FT  TRANSMEM 45 65 POTENTIAL.
FT  TRANSMEM 79 99 POTENTIAL.
FT  TRANSMEM 102 122 POTENTIAL.
FT  TRANSMEM 131 151 POTENTIAL.
FT  TRANSMEM 157 177 POTENTIAL.
FT  TRANSMEM 204 224 POTENTIAL.
FT  TRANSMEM 235 255 POTENTIAL.
FT  TRANSMEM 270 290 POTENTIAL.
FT  TRANSMEM 291 311 POTENTIAL.
FT  TRANSMEM 325 345 POTENTIAL.
FT  TRANSMEM 348 368 POTENTIAL.
FT  CONFLICT 265 265 L -> H (IN REF. 2).
SQ  SEQUENCE 382 AA; 41431 MW; 9C0F47A928B7C949 CRC64;

Query Match 66.7%; Score 30; DB 1; Length 382;
Best Local Similarity 60.0%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
DB 277 LGSIAMLSQA 286
:|||||:
ID DCTS_RHOCA STANDARD; PRT; 657 AA.
AC F37739;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE C4-DICARBOXYLATE TRANSPORT SENSOR PROTEIN DCTS (EC 2.7.3.-).
GN DCTS.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B10;
RX MEDLINE-93204897; PubMed-8455557;
RA Hamlin M.J., Shaw J.G., Kelly D.J.;
RT "Sequence analysis and interposon mutagenesis of a sensor-kinase
    (Dcts) and response-regulator (Dctr) controlling synthesis of the
    high-affinity C4-dicarboxylate transport system in Rhodobacter
```

```

capsulatus.";
Mol. Gen. Genet. 237:215-224(1993).
-1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM DCTS/DCTR
    INVOLVED IN THE TRANSPORT OF C4-DICARBOXYLATES. DCTS FUNCTIONS AS
    A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT PHOSPHORYLATES DCTR IN
    RESPONSE TO ENVIRONMENTAL SIGNALS.
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
-1- SIMILARITY: TO OTHER PROKARYOTIC SENSORY TRANSDUCTION HISTIDINE
    KINASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL; X64733; CAA45999.1; -
DR  PIR; S30288; S30288.
DR  InterPro; IPR000014; -.
DR  InterPro; IPR000410; -.
DR  InterPro; IPR001610; -.
DR  Pfam; PF00785; PAC; 1.
DR  Pfam; PF00989; PAS; 1.
DR  Pfam; PF00512; signal; 1.
DR  PRINTS; PR00344; BCTRLSENSOR.
KW Sensory transduction; Transferase; Kinase; Phosphorylation;
KW Transmembrane; Inner membrane.
FT  DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM 27 51 POTENTIAL.
FT  DOMAIN 52 252 PERIPLASMIC (POTENTIAL).
FT  TRANSMEM 253 273 POTENTIAL.
FT  DOMAIN 274 657 CYTOPLASMIC (POTENTIAL).
FT  DOMAIN 407 422 INTER-DOMAIN LINKER (POTENTIAL).
FT  MOD_RES 440 440 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ  SEQUENCE 657 AA; 70142 MW; 139D2CFC7CFFA69B CRC64;

Query Match 66.7%; Score 30; DB 1; Length 657;
Best Local Similarity 40.0%; Pred. No. 48;
Matches 4; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
DB 399 MGSVIDITQA 408
|||||:
ID RPAI_SULAC STANDARD; PRT; 880 AA.
AC P11512;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE SUBUNIT A' (EC 2.7.7.6).
GN RPOA1 OR RPOA.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE-89315197; PubMed-2501756;
RA Puchler G., Lottspeich F., Zillig W.;
RT "Organization and nucleotide sequence of the genes encoding the large
    subunits A, B and C of the DNA-dependent RNA polymerase of the
    archaebacterium Sulfolobus acidocaldarius.";
RL Nucleic Acids Res. 17:4517-4534(1989).
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
    OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
    SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
```

```

-!- SIMILARITY: SOME, WITH SUCROSE SYNTHASES. -----
This SWISS-prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement agreement remains unaltered in all commercial entities requiring a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licens@isb-sib.ch) -----

EMBL; Z56378; CAA91217.1; -
EMBL; Z48640; CAA88587.1; -
InterPro; IPR001296; -
Pfam; PF00534; Glycosyltransferase_1_phosphorylation.
Sequence 1059 AA, 118203 MW, FE3B49081A48EC99 CRC64;

Query Match      66.7%; Score 30; DB 1; Length 1059;
Best Local Similarity 50.0%; Pred. No. 82;
Matches          5; Conservative    4; Mismatches   1; Indels     0; Gaps      0;

OY      1 MGSIVLSOS 10
        |||l||:::
Db       286 MGHIIMSKA 295

RESULT 15
EFXX YEAST YEAST STANDARD; PRT; 1277 AA.
AD Q03490
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 135 KDA SUBUNIT (EIF3 P135)
DE (TRANSLATION INITIATION FACTOR EIF3, P135 SUBUNIT).
DE TIF31 OR CLU1 OR ERMO12W OR Baked's Yeast).
GN Sakayotai; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Eukaryota; Fungi; Ascomycota; Saccharomycetaceae; Saccharomycetes.
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxId=4932;

SEQUENCE FROM N.A.
STRAIN=S288C AB972;
Vornlocher H. P.; Hanachi P., Ribeiro S., Hershey J.W.B.:
"A 110-kilodalton subunit of translation initiation factor eif3 and an associated 135-kilodalton protein are encoded by the Saccharomyces cerevisiae TIF32 and TIF31 genes."
J. Biol. Chem. 274:16802-16812(1999).

(2) SEQUENCE FROM N.A.
STRAIN=S288C AB972;
Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF METHIONYL-TRNAI AND MRNA.
-!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 8 DIFFERENT SUBUNITS.
-!- SIMILARITY: BELONGS TO THE EIF-3 P135 FAMILY. -----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licens@isb-sib.ch) -----

EMBL; AF004911; AAB82415.1; -
EMBL; Z48613; CAAB8529.1; -
SGD; S0004614; CLU1.
Initiation factor; Protein biosynthesis.
SEQUENCE 1277 AA; 145165 MW; 1E643F4EE7ED5B5 CRC64;

```

Db 710 LGKIELSO 718

RESULT	17				
YP23_STA00					
ID	YP23_STA00	STANDARD;	PRT;	188	AA.
AC	P3217;				
DT	01-NOV-1991	(Rel. 20, Created)			
DT	01-NOV-1991	(Rel. 20, Last sequence update)			
DT	01-NOV-1995	(Rel. 32, Last annotation update)			
DE	HYPOTHETICAL TRANSCRIPTIONAL REGULATOR IN QACA 5' REGION (ORF 188).				
OS	Staphylococcus aureus.				
OC	Plasmid pSKI.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;				
OC	Bacillus/Staphylococcus group; Staphylococcus.				
OX	NCBI_TaxID=1280;				
RN	[1]				

SEQUENCE FROM N.A.
RX MEDLINE=91211614; PubMed=2089219;
RA Rouch D.A., Cram D.S., Diberardino D., Littlejohn T.G.,
RA Skurray R.A.;
RT "Efflux-mediated antiseptic resistance gene qacA from *Staphylococcus*
RT aureus: common ancestry with tetracycline- and sugar-transport
RT proteins.";
RT Mol. Microbiol. 4:2051-2062(1990).
RL
CC -I- FUNCTION: POSSIBLE TRANSCRIPTIONAL REPRESSOR OF THE *QACA* GENE.
CC -I- SIMILARITY: BELONGS TO THE TETR/ACRS FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

 EMBL; X56628; CAA39962.1; -.
 PIR; S12393; S12393.
 InterPro; IPR001647; -.
 Pfam; PF00440; tetr; 1.
 PRINTS; PR00455; HTHTETR.
 PROSITE; PS01081; HTH_TETR_FAMILY; 1.
 Hypothetical protein; Transcription regulation; DNA-binding; Plasmid;
 Repressor.
 24 43 H-T-H MOTIF (POTENTIAL).
 DNA_BIND 188 AA; 22174 MW; 7B91E005C8047322 CRC64;
 SO SEQUENCE

Query Match	64.4%	Score 29;	DB 1;	Length 188;
Best Local Similarity	66.7%	Pred. No. 20;		
Matches	6:	Conservative	2;	Mismatches
				Gaps 0;
				Indels 1;

GN RV3679 OR MIV023.027.
OS *Mycobacterium tuberculosis*.

DR	PIR:	S0430047	S049004	
DR	HSP:	P24295:	1ADP:	
DR	interPro:	IPR001625:		
DR	PRINTS:	P000082:	GLFDIDRGNASE	1
DR	PROSITE:	PS00074:	GLFV_DEHYDROGENASE:	1
KW	Oxidoreductase:	NADP:		
ACT_SITE		414	114	
FT	SEQUENCE	435 AA:	49608 MW:	6823964395C002B7D CRC64:
SQ				

Query Match 64.4%; Score 29; DB 1; Length 459;
Best Local Similarity 66.7%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSVQLSQS 10
II: I I I I
DB 245 GSVVLSLDS 253

RESULT 21
SELA_ECOLI
ID SELA_ECOLI STANDARD; PRT; 463 AA.
AC P23328; P78119;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1993 (Rel. 38, Last annotation update)
DE L-SERYL-TRNA(SEC) SELENIUM TRANSFERASE (EC 2.9.1.1) (CYSTEINYL-
TRNA(SEC) SELENIUM TRANSFERASE) (SELENOCYSTEINE SYNTHASE)
DE (SELENOCYSTEINYL-TRNA(SEC) SYNTHASE).
GN SELA OR FDHA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91177883; PubMed=2007584;
RA Forchhammer K., Leinfelder W., Boesmler K., Veprek B., Boeck A.;
RT "Selenocysteine synthase from Escherichia coli. Nucleotide sequence
of the gene (selA) and purification of the protein.";
RL J. Biol. Chem. 266:6318-6323(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -!- FUNCTION: CONVERTS SERYL-TRNA-UCA TO SELENOCYSTEINYL-TRNA-UCA
DURING SELENOPROTEIN BIOSYNTHESIS.
CC -!- CATALYTIC ACTIVITY: L-SERYL-TRNA(SEC) + SELENOPHOSPHATE =
L-SELENOCYSTEINYL-TRNA(SEC) + H(2)O + ORTHOPHOSPHATE
CC -!- COFACTOR: PYRIDOXAL PHOSPHATE. THIS REACTION REQUIRES ATP,
MAGNESIUM AND REDUCED SELENIUM.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M64177; AAA24624.1; -;
DR EMBL; U00039; AAB18568.1; -;
DR EMBL; AE000436; AAC76615.1; -;
DR PIR; A38730; A38730.
DR ECO2DBASE; G050.7; 6TH EDITION.
DR EcoGene; Egl0941; selA.
KW Transferase; Pyridoxal phosphate; Magnesium; Selenium.
FT CONFLICT 7 F -> S (IN REF. 2).
SQ SEQUENCE 463 AA; 50667 MW; A2AC08A84E8ECD01 CRC64;

Query Match 64.4%; Score 29; DB 1; Length 463;
Best Local Similarity 75.0%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSVQLSQS 9
II: I I I I
DB 262 GSVLDSQ 269

RESULT 22
VGLY_PIARY
ID VGLY_PIARY STANDARD; PRT; 503 AA.
AC P03540;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE GLYCOPROTEIN POLYPROTEIN PRECURSOR [CONTAINS: GLYCOPROTEINS G1 AND
G2].
DE G2].
GN GPC.
OS Pichinde arenavirus.
OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.
OX NCBI_TaxID=11630;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85033957; PubMed=6492264;
RA Auperin D.D., Romanowski V., Galinski M., Bishop D.H.L.;
RT "Sequencing studies of pichinde arenavirus S RNA indicate a novel
coding strategy, an ambisense viral S RNA.";
RL J. Virol. 52:897-904(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87160943; PubMed=2435460;
RA Bishop D.H.L., Auperin D.D.;
RT "Arenavirus gene structure and organization";
RL Curr. Top. Microbiol. Immunol. 133:5-17(1987).
CC -!- SIMILARITY: BELONGS TO THE ARENAVIRUSES GPC PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02734; AAA46824.1; -;
DR EMBL; M16735; AAA46827.1; -;
DR PIR; A04149; Q0XPGP.
DR InterPro; IPR001535; -;
DR Pfam; PF00798; Arena glycoprot; 1.
KW Polyprotein; Glycoprotein; Envelope protein.
FT CHAIN 1 271 GLYCOPROTEIN G1.
FT CHAIN 272 503 GLYCOPROTEIN G2.
FT CARBOHYD 67 67 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 111 111 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 132 132 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 241 241 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 379 379 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 404 404 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 503 AA; 57278 MW; 17740E092B450044 CRC64;

Query Match 64.4%; Score 29; DB 1; Length 503;
Best Local Similarity 70.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
II: I I I I
DB 1 MGQIVTLIQS 10

Query Match 64.4%; Score 29; DB 1; Length 540;
Best Local Similarity 60.0%; Pred. No. 66;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVOLSOS 10
||| :|||

Db 497 MGSIDEMSOS 506

RESULT 25
SKN7_YEAST STANDARD; PRT; 622 AA.

AC P38889; P39747;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR SKN7 (POS9 PROTEIN).
GN SKN7 OR POS9 OR BRY1 OR YH206W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94042854; PubMed=8226633;
RA Brown J.L., North S., Bussey H.;
RT "SKN7, a yeast multicopy suppressor of a mutation affecting cell wall
beta-glucan assembly, encodes a product with domains homologous to
prokaryotic two-component regulators and to heat shock transcription
factors.";
RL J. Bacteriol. 175:6908-6915(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Krens B., Charizanis C., Entian K.-D.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du C., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII.";
RL Science 265:2077-2082(1994).
RN [4]
RP FUNCTION, AND MUTAGENESIS.
RX MEDLINE=95045411; PubMed=7957083;
RA Brown J.L., Bussey H., Stewart R.C.;
RT "Yeast Skn7p functions in a eukaryotic two-component regulatory
pathway.";
RL EMBO J. 13:5186-5194(1994).
CC -!- FUNCTION: INVOLVED IN OXIDATIVE STRESS. TRANSCRIPTION FACTOR THAT
MAY FUNCTION IN A TWO-COMPONENT SIGNAL TRANSDUCTION PATHWAY THAT
ACTS IN PARALLEL WITH THE PKC1 CASCADE TO REGULATE GROWTH AT THE
CELL SURFACE.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE HSF FAMILY.
CC -!- SIMILARITY: TO PROKARYOTE SENSORY TRANSDUCTION PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: U00485; AAC48911.1; -;
DR EMBL: X83031; CAA58143.1; -;
DR EMBL: U00029; AAB69734.1; -;

DR PIR: A49344; A49344.
DR PIR: S48987; S48987.
DR HSSP: P22121; 2HTS.
DR SGD: S0001249; SKN7.
DR InterPro: IPR000232; -.
DR InterPro: IPR001789; -.
DR Pfam: PF00447; HSF_DNA-bind; 1.
DR Pfam: PF00072; response_reg; 1.
DR PRINTS: PR00056; HSFDOMAIN.
DR PROSITE: PS00434; HSF_DOMAIN; 1.
KW Transcription regulation; Sensory transduction; Nuclear protein;
KW DNA-binding; Phosphorylation.
FT DNA_BIND 86 190
FT MOD_RES 427 427 PHOSPHORYLATION (PROBABLE).
FT MUTAGEN 427 427 D->N: DIMINISHED ACTIVITY.
FT MUTAGEN 427 427 D->E: AUGMENTED ACTIVITY.
SQ SEQUENCE 622 AA; 69202 MW; 4C732FD66E326742 CRC64;

Query Match 64.4%; Score 29; DB 1; Length 622;
Best Local Similarity 70.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGSIVOLSOS 10
||| |||

Db 605 MGSTPQLPQS 614

RESULT 26
DIA3_MOUSE STANDARD; PRT; 1171 AA.

AC Q92207;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIAPHANOUS PROTEIN HOMOLOG 3 (DIAPHANOUS-RELATED FORMIN 3) (DRF3)
DE (MDIA2) (P134MDIA2).
GN DIAPH3 OR DIAP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98204843; PubMed=9535835;
RA Alberts A.S., Bouquin N., Johnston L.H., Treisman R.;
RT "Analysis of RhoA-binding proteins reveals an interaction domain
conserved in heterotrimeric G protein beta subunits and the yeast
response regulator protein Skn7.";
RL J. Biol. Chem. 273:8616-8622(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Tomimaga T., Sahai E., Treisman R.H., Alberts A.S.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP FUNCTION.
RX MEDLINE=20142655; PubMed=10678165;
RA Tomimaga T., Sahai E., Chardin P., McCormick F., Courtneidge S.A.,
RA Alberts A.S.;
RT "Diaphanous-related formins bridge Rho GTPase and Src tyrosine kinase
signaling.";
RL Mol. Cell 5:13-25(2000).

CC -!- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
SERUM RESPONSE FACTOR. DRF PROTEINS COUPLE RHO AND SRC TYROSINE
KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS.
CC -!- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION.
CC -!- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FH1) DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 2 (FH2) DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOGY 3 (FH3) DOMAIN.
 CC -2- SIMILARITY: CONTAINS 1 THE NO REGULATORY DOMAIN (DND).
 CC -3- SIMILARITY: BELONGS TO THE FORMIN HOMOLOGY FAMILY. DIAPHANOUS
 CC SUBFAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no
 CC way modified and the source is acknowledged. For commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF094519; AAC71171.1;
 CC MDG; MG1:1927222; Diap3.
 CC DR InterPro: IPR000304; -
 CC DR GBD.
 CC KW Coiled coil, Repeat.
 CC FT DOMAIN 81 277
 CC FT DOMAIN 176 473
 CC FT DOMAIN 373 403
 CC FT DOMAIN 478 533
 CC FT DOMAIN 540 610
 CC FT DOMAIN 610 716
 CC FT DOMAIN 887 1038
 CC FT DOMAIN 988 1038
 CC FT DOMAIN 1039 1053
 CC FT DOMAIN 1055 1058
 CC FT SEQUENCE 1171 AA; 133685 MW; 95347A854CABC7CF CRC64;
 CC
 CC Query Match 64.4%; Score 29; DB 1; Length 1171;
 CC Best Local Similarity 60.0%; Pred. No. 1.6e+02; Indels 0; Gaps 0;
 CC Matches 6; Conservative 2; Mismatches 2;
 CC
 CC Qy 1 MGSIVQLSQS 10
 CC Db 950 LGSMTQLVQS 959
 CC
 CC RESULT 27
 CC CLH2_HUMAN STANDARD; PRT; 1640 AA.
 AC P53675; Q14017; Q15808; Q15809;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-2000 (Rel. 45, Last sequence update)
 DE CLH2_HUMAN
 DE CLATHRIN HEAVY CHAIN 2 (CLH-22)
 GN CLH2 OR CLND OR CLH22 OR CLTCL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN MIM: 601273;
 RC TISSUE=Fetal brain;
 RA MEDLINE=963111556; PubMed=8733128;
 RA Sirokin H., Morrow B., Dasgupta R., Goldberg R., Patangali S.R.,
 RA Shi G., Cannizzaro L., Shprintzen R., Weissman S., Kucherlapati R.,
 RA "Isolation of a new clathrin heavy chain gene with muscle-specific
 RA expression, from the region commonly deleted in velo-cardio-facial
 RA syndrome". Hum. Mol. Genet. 5:617-624(1996).
 RN [2]
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RC TISSUE=Fetal brain, and skeletal muscle;
 RX MEDLINE=963111557; PubMed=8733129;
 RA Kedra D., Peyrard M., Fransson I., Collins J.E., Dunham I.,
 RA "Characterization of a second human clathrin heavy chain polypeptide
 RA gene (CLH-22) from chromosome 22q11.";
 RL Hum. Mol. Genet. 5:625-631(1996).
 RN [3]

RP SEQUENCE FROM N.A. PubMed=8841170;
 RA MEDLINE=97061171;
 RA "Cloning and characterization of a novel human clathrin heavy chain
 RT gene (CLTCL).";
 RT Genomics 35:466-472(1996).
 RL
 CC -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
 CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES
 CC ARE INVOLVED IN THE TRIMERIZATION OF THE PLASMA MEMBRANE OR TO THE
 CC TRANS GOLGI NETWORK (BY SIMILARITY).
 CC -1- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
 CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE
 CC PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE
 CC PH AND THE CONCENTRATION OF CALCIUM (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
 CC VESICLES (BY SIMILARITY).
 CC -1- MEMBRANE PRODUCTION: 2 ISOFORMS: A LONG/BRAIN FORM (SHOWN HERE)
 CC AND A SHORT/MUSCLE FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: MAXIMAL LEVELS IN SKELETAL MUSCLE. HIGH LEVELS
 CC IN HEART AND TESTIS. LOW EXPRESSION DETECTED IN ALL OTHER TISSUES.
 CC -1- DOMAIN: THE C-TERMINAL THIRD OF THE HEAVY CHAINS FORMS THE HUB OF
 CC THE TRISKELION. THIS REGION CONTAINS THE TRIMERIZATION DOMAIN AND
 CC THE LIGHT-CHAIN BINDING DOMAIN INVOLVED IN THE ASSEMBLY OF THE
 CC CLATHRIN COAT.
 CC -1- DISEASE MAY PLAY A ROLE IN THE HYPOTONIA SEEN IN VELO-CARDIO-
 CC -1- FACIAL SYNDROME (VCFS).
 CC -1- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U41763; AAC50494.1; -
 CC EMBL; X95489; CAA64753.1; -
 CC EMBL; X65489; CAA64753.1; -
 CC EMBL; U60802; AAB40908.1; -
 CC EMBL; U60803; AAB40909.1; -
 CC MIM: 601273; -
 CC InterPro: IPR000547; -
 CC InterPro: IPR001473; -
 CC Pfam: PF01394; Clathrin_Protein; 7.
 CC Pfam: PF00637; Clathrin_Protein; 7.
 CC Coated pits; Alternative splicing.
 CC DOMAIN 1 479 GLOBULAR TERMINAL DOMAIN.
 CC DOMAIN 480 523 FLEXIBLE LINKER.
 CC DOMAIN 524 1640 HEAVY CHAIN ARM.
 CC DOMAIN 524 634 DISTAL SEGMENT.
 CC DOMAIN 639 1640 PROXIMAL SEGMENT.
 CC DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,
 CC (POTENTIAL) LATTICE DISASSEMBLY
 CC BINDING 1213 1522 LIGHT CHAIN (BY SIMILARITY).
 CC DOMAIN 1551 1640 TRIMERIZATION (BY SIMILARITY).
 CC VARSPIC 1479 1535 MISSING (IN SHORT ISOFORM).
 CC CONFLICT 193 193 P -> H (IN REF. 1).
 CC CONFLICT 215 215 L -> H (IN REF. 1).
 CC CONFLICT 215 215 K -> H (IN REF. 2).
 CC CONFLICT 530 530 K -> Q (IN REF. 1).
 CC CONFLICT 530 530 E -> K (IN REF. 1).
 CC CONFLICT 691 691 E -> K (IN REF. 3).
 CC CONFLICT 1474 1474 E -> K (IN REF. 3).
 CC CONFLICT 1620 1640 RKQEEHTEPAPLVDFDGHGHE -> PPSKRSK (IN
 CC REF. 3).
 CC SEQUENCE 1640 AA; 187029 MW; C661E1AB989D8E7F CRC64;
 CC
 CC Query Match 64.4%; Score 29; DB 1; Length 1640;
 CC Best Local Similarity 66.7%; Pred. No. 2.3e+02;
 CC Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MGSIVQLSQ 9
DB 715 LGSIVNFSQ 723

RESULT 28
ID CLH1_HUMAN STANDARD; PRT; 1675 AA.
AC Q00610;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CLATHRIN HEAVY CHAIN 1 (CLH-17).
GN CLH1 OR CLH17 OR KIAA0034.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1].
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051387; PubMed=7584026;
RA Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayashi Y.,
RA Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. I.
RT The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT analysis of randomly sampled cDNA clones from human immature myeloid
RT cell line KG-1."
RL DNA Res. 1:27-35(1994).
RN [2].
RP SEQUENCE OF 560-864 FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=92112210; PubMed=1765375;
RA Stokes D.G., Iozzo R.V.;
RT "Human clathrin heavy chain (CLTC): partial molecular cloning,
RT expression, and mapping of the gene to human chromosome 17q11-qter."
RL Genomics 11:174-178(1991).
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES
CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE
CC TRANS GOLGI NETWORK.
CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE
CC PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE
CC PH AND THE CONCENTRATION OF CALCIUM.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC VESICLES.
CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D21260; BAA04801.1; -.
CC EMBL; X55878; CAA39363.1; -.
CC PIR; A40573; A40573.
CC MIM; 118955; -.
CC InterPro; IPR000547; -.
CC InterPro; IPR001473; -.
CC Pfam; PF01394; Clathrin_propel; 7.
CC Pfam; PF00637; Clathrin_repeat; 7.
CC Coated pits.
CC Coated pits. 1 479 GLOBULAR TERMINAL DOMAIN.
CC DOMAIN 480 523 FLEXIBLE LINKER.
CC DOMAIN 524 1675 HEAVY CHAIN ARM.
CC DOMAIN 524 634 DISTAL SEGMENT.
CC DOMAIN 639 1675 PROXIMAL SEGMENT.
CC DOMAIN 449 465 BINDING SITE FOR THE UNCOATING ATPASE,

INVOLVED IN LATTICE DISASSEMBLY
(POTENTIAL).
LIGHT CHAIN (BY SIMILARITY).
TRIMERIZATION (BY SIMILARITY).
Q -> R (IN REF. 2).
FT BINDING 1213 1522
FT DOMAIN 1550 1675
FT CONFLICT 560 560
FT CONFLICT 817 817
SQ SEQUENCE 1675 AA; 191614 MW; 6C4F2D54950079E2 CRC64;

Query Match 64.4%; Score 29; DB 1; Length 1675;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
DB 715 LGSIVNFSQ 723

RESULT 29
ID CLH_BOVIN STANDARD; PRT; 1675 AA.
AC P49951;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CLATHRIN HEAVY CHAIN.
GN CLTC.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1].
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=96028100; PubMed=7585943;
RA Liu S.-H., Wong M.L., Craik C.S., Brodsky F.M.;
RT "Regulation of clathrin assembly and trimerization defined using
RT recombinant triskelion hubs."
RL Cell 83:257-267(1995).
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC COATED PITS & VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES
CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE
CC TRANS GOLGI NETWORK.
CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE
CC PRESENCE OF LIGHT CHAINS, HUB ASSEMBLY IS INFLUENCED BY BOTH THE
CC PH AND THE CONCENTRATION OF CALCIUM.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC VESICLES.
CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U31757; AAC48524.1; -.
CC InterPro; IPR000547; -.
CC InterPro; IPR001473; -.
CC Pfam; PF01394; Clathrin_propel; 7.
CC Pfam; PF00637; Clathrin_repeat; 7.
CC Coated pits.
CC Coated pits. 1 479 GLOBULAR TERMINAL DOMAIN.
CC DOMAIN 480 523 FLEXIBLE LINKER.
CC DOMAIN 524 1675 HEAVY CHAIN ARM.

```

```

FT DOMAIN 524 634 DISTAL SEGMENT.
FT DOMAIN 634 1675 BINDING SITE FOR THE UNCOATING ATPASE,
FT DOMAIN 449 465 INVOLVED IN LATTICE DISASSEMBLY
(POTENTIAL).
FT BINDING 1213 1522 LIGHT CHAIN.
FT DOMAIN 1550 1675 TRIMERIZATION.
FT SEQUENCE 1675 AA; 191587 MW; 6CAF2D54801579E2 CRC64;
SQ

Query Match 64.4%; Score 29; DB 1; Length 1675;
Best Local Similarity 66.7%; Pred. No. 2.3e+02; Indels 0; Gaps 0;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVOLSQ 9
DB 715 LGSIVNFSQ 723

RESULT 30
CLH_RAT STANDARD; PRT: 1675 AA.
ID CLH_RAT
AC P34574;
DT 01-FEB-1994 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE CLATHRIN HEAVY CHAIN.
GN CLTC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Euteleostei; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_taxid=10116;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=88097376; PubMed=3480512;
RA Kirchhausen T., Harrison S.C., Chow E.P., Mattaliano R.J.,
RA Ramchandran K.L., Smart J., Brosius J.;
RA Clathrin heavy chain: molecular cloning and complete primary
RA structure determined.
RL Proc. Natl. Acad. Sci. U.S.A. 84:8805-8809(1987).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-493.
RX MEDLINE=99043510; PubMed=9827808;
RA Ter Haar E., Musacchio A., Harrison S.C., Kirchhausen T.;
RA Atomic structure of clathrin: a beta propeller terminal domain joins
RA the clathrin cage.
RL Cell 95:553-573(1998)
CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
CC COATED PITS AND VESICLES. TWO DIFFERENT ADAPTOR PROTEIN COMPLEXES
CC LINK THE CLATHRIN LATTICE EITHER TO THE PLASMA MEMBRANE OR TO THE
CC TRANS GOLGI NETWORK.
CC -!- SUBUNIT: CLATHRIN TRISKELIONS, COMPOSED OF 3 HEAVY CHAINS AND 3
CC LIGHT CHAINS, ARE THE BASIC SUBUNITS OF THE CLATHRIN COAT. IN THE
CC PRESENCE OF ADAPTOR PROTEIN COMPLEXES, THE COAT IS INFLUENCED BY BOTH THE
CC PH AND THE CONCENTRATION OF CALCIUM.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
CC VESICLES.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO THE CLATHRIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z30423; AAA00874.1;
CC PIR; S42369; S42369.
CC WormPep; T20G5.1; CE00480.
CC InterPro; IPR000547;
CC DR PDB; 1BPO; 16-APR-99.
CC InterPro; IPR000547;
CC DR InterPro; IPR001473;
CC KW Hypothetical protein; Coated pits.
CC SEQUENCE 1681 AA; 191541 MW; 44D15C61339009D9 CRC64;

```

```
Query Match          64.4%; Score 29; DB 1; Length 1681;
Best Local Similarity 66.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
    :|||||
DB 717 LGSIVNFSQ 725

RESULT 32
TELL_YEAST
ID TELL_YEAST STANDARD; PRT; 2787 AA.
AC P38110;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TELOMER LENGTH REGULATION PROTEIN TELL.
GN TELL OR YBL088C OR YBL0706.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96076635; PubMed=7502586;
RA Obermayer B., Gassenhuber J., Piravandi E., Domdey H.;
RT "Sequence analysis of a 78.6 kb segment of the left end of
RT Saccharomyces cerevisiae chromosome II.";
RL Yeast 11:1103-1112(1995).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RA Greenwell P.S., Krommal S.L., Porter S.E., Gassenhuber J.,
RA Obermayer B., Petes T.D.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DDJB databases.
CC -!- FUNCTION: PUTATIVE PHOSPHATIDYLINOSITOL KINASE INVOLVED IN
CC CONTROLLING TELOMERE LENGTH.
CC -!- SIMILARITY: BELONGS TO THE P13/P14-KINASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79489; CAA56016.1; -
DR EMBL; Z35849; CAA84909.1; -
DR EMBL; U31331; AAB69802.1; -
DR PIR; S45416; S45416.
DR SGD; S0000184; TELL.
DR InterPro; IPR000403; -
DR Pfam; PF00454; P13_P14_Kinase; 1.
DR PROSITE; PS00915; P13_4_KINASE_1; 1.
DR PROSITE; PS00916; P13_4_KINASE_2; 1.
DR PROSITE; PS0290; P13_4_KINASE_3; 1.
KW Transferase; Kinase.
FT DOMAIN 2461 PI3K/PI4K.
FT CONFLICT 1190 1190 Y -> F (IN REF. 2).
SQ SEQUENCE 2787 AA; 321663 MW; 439B6E189E39499B CRC64;

Query Match          64.4%; Score 29; DB 1; Length 2787;
Best Local Similarity 62.5%; Pred. No. 4.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSIVQLS 8
    :|||||
DB 1996 LGSIIOLA 2003
```

```
RESULT 33
G3P_MYCGE
ID G3P_MYCGE STANDARD; PRT; 337 AA.
AC P47543;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE (EC 1.2.1.12) (GAPDH).
GN GAPA-OR GAP OR MG301.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier J.C.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 1-81 AND 279-337 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bitt K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE + ORTHOPHOSPHATE
CC + NAD(+) = 1,3-DIPHOSPHATEGLYCERATE + NADH.
CC -!- PATHWAY: FIRST STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39710; AAC71523.1; -
DR EMBL; U02213; AAD12507.1; -
DR EMBL; U02178; AAD12463.1; -
DR DR HSSP; P17721; 1HDG.
DR TIGR; MG301; -
DR InterPro; IPR000173; -
DR Pfam; PF00044; gpdh; 1.
DR PRINTS; PR00078; G3PDHGRNASE.
DR PROSITE; PS00071; GAPDH; 1.
KW Glycolysis; Oxidoreductase; NAD.
FT BINDING 157 157 GLYCERALDEHYDE 3-PHOSPHATE
FT (BY SIMILARITY).
FT ACT_SITE 184 184 ACTIVATES THIOL GROUP DURING CATALYSIS
FT (BY SIMILARITY).
SQ SEQUENCE 337 AA; 37097 MW; FALEA1966687006B CRC64;

Query Match          62.2%; Score 28; DB 1; Length 337;
Best Local Similarity 85.7%; Pred. No. 66;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 GSIVQLS 8
    :|||||
```

```

OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
OO [1]
SEQUENCE FROM N.A. PubMed-7798319;
MEDLINE=93056177;
PMID=7798319;
BYpassing anaphase by fission yeast cut9 mutation: requirement of
cut9+ to initiate anaphase."
J. Cell Biol. 127:1655-1670(1994).
!- FUNCTION: INTERACTS WITH CUT9.
!!- SIMILARITY: TO YEAST YMR262W.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licensedis@isb-sib.ch)
CC -----
CC EMBL: D31845; BAA06631.1;
CC InterPro: IPR001130;
CC Pfam: PF01026; UPE0006; 1.
CC SEQUENCE 387 AA; 44421 MW; AAC154B805BC1085 CRC64;
QY Query Match 92.2%; Score 28; DB 1; Length 387;
DB Best Local Similarity 92.0%; Prev % 75%
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 GSIVQLSQ 9
DB 294 GSIEQISQ 301
||| |||
RESULT 36
COAT_BBV STANDARD; PRT; 407 AA.
AC P04329;
DC 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE COAT PROTEIN PRECURSOR.
DE Black beetle virus (BBV).
OS Black beetle virus (BBV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nodaviridae;
OC Alphonaviruses.
OX NCBI_TaxID=12285;
[1] SEQUENCE FROM N.A. PubMed=6548308;
MEDLINE=8501452; PMID=6548308;
Quarino L.A., Kaesberg P.;
"Primary and secondary structure of black beetle virus RNA2, the
genomic messenger for BBV coat protein precursor.";
Nucleic Acids Res. 12:7215-7223(1984).
[2] X-RAY CRYSTALLOGRAPHY, AND SIMILARITY TO OTHER NODAVIRUSES.
X-RAY CRYSTALLOGRAPHY, AND SIMILARITY TO OTHER NODAVIRUSES.
MEDLINE=90339486; PubMed=2116525;
Kaesberg P., Dasgupta R., Sgro J.-Y., Wery J.-P., Selling B.H.,
"Structural homology among four nodaviruses as deduced by sequencing
and x-ray crystallography.";
J. Mol. Biol. 214:423-435(1990).
[3] X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
MEDLINE=94118310; PubMed=8289282;
Wery J., Reddy S., Goswami, Johnson J.E.;
"Refined three-dimensional structure of an insect virus at 2.8-A
resolution".
J. Mol. Biol. 235:565-586(1994).
```

CC -!- PTM: ENZYMATICALLY CLEAVED INTO COAT PROTEINS BETA AND GAMMA.
CC HOWEVER, THE EXACT CLEAVAGE SITE HAS NOT BEEN DETERMINED.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X00956; CAA25468.1;
CC PIR: A04151; VCB22G.
CC PIR: S11036; S11036.
CC PDB: 2BBV; 31-AUG-94.
CC MEROPS: A06.001; -.
CC InterPro: IPR000696; -.
CC Pfam: PF01829; Peptidase A6; 1.
CC PRINTS: PR00863; NODAVIRPTASE.
KW Coat protein: Hydrolase; Aspartyl protease; 3D-structure.
SQ SEQUENCE 407 AA; 4388 MW; 3A6C6C9A98A5C26C CRC64;

Query Match 62.2%; Score 28; DB 1; Length 407;
Best Local Similarity 55.6%; Pred. No. 82;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
| : : : |
Db 53 MGALTRLQ 61

RESULT 37
DCTA_ECOLI
ID DCTA_ECOLI STANDARD; PRT; 428 AA.
AC P37312;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE C4-DICARBOXYLATE TRANSPORT PROTEIN.
GN DCTA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-94316500; PubMed-8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -!- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF DICARBOXYLATES SUCH AS
CC SUCCINATE, FUMARATE, AND MALATE FROM THE PERIPLASM ACROSS THE
CC INNER MEMBRANE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE SODIUM:DICARBOXYLATE SYMPORTER FAMILY
CC (SDF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00039; AAB18505.1;
CC EMBL: AE000429; AAC76553.1;
CC EcoGene: EG20044; dcta.
CC InterPro: IPR001991; -.

DR Pfam: PF00375; SDF; 1.
DR PROSITE: PS00713; NA_DICARBOXYL_SYM_1; 1.
DR PROSITE: PS00714; NA_DICARBOXYL_SYM_2; 1.
KW Transport; Transmembrane; Inner membrane; Symport.
FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 21 39 1 (POTENTIAL).
FT DOMAIN 40 58 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 59 77 2 (POTENTIAL).
FT DOMAIN 78 89 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 90 108 3 (POTENTIAL).
FT DOMAIN 109 139 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 140 158 4 (POTENTIAL).
FT DOMAIN 159 161 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 162 180 5 (POTENTIAL).
FT DOMAIN 181 199 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 200 218 6 (POTENTIAL).
FT DOMAIN 219 231 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 232 250 7 (POTENTIAL).
FT DOMAIN 251 271 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 272 290 8 (POTENTIAL).
FT DOMAIN 291 310 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 311 329 9 (POTENTIAL).
FT DOMAIN 330 340 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 341 359 10 (POTENTIAL).
FT DOMAIN 360 368 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 369 388 11 (POTENTIAL).
FT DOMAIN 389 395 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 396 414 12 (POTENTIAL).
FT DOMAIN 415 428 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 428 AA; 45436 MW; D9B32F987B62D234 CRC64;

Query Match 62.2%; Score 28; DB 1; Length 428;
Best Local Similarity 55.6%; Pred. No. 86;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSIVQLSQ 9
| : : : |
Db 213 VGLVQLGQ 221

RESULT 38
SMVA_SALTY
ID SMVA_SALTY STANDARD; PRT; 496 AA.
AC P37594;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE METHYL VIOLOGEN RESISTANCE PROTEIN SMVA.
GN SMVA.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SL1303;
RX MEDLINE-95011654; PubMed-7926834;
RA Hongo E., Morimyo M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,
RA Ichimura S., Noda Y.;
RT "The methyl viologen-resistance-encoding gene smva of Salmonella
RT typhimurium."
RL Gene 148:173-174(1994).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC EMBL: D26057; BAA05055.1; -
 DR StyGene; SG10384; smva.
 DR Transmembrane; Inner membrane.
 KW TransmemEM 5 25 POTENTIAL.
 FT TransmemEM 44 64 POTENTIAL.
 FT TransmemEM 92 93 POTENTIAL.
 FT TransmemEM 102 115 POTENTIAL.
 FT TransmemEM 135 155 POTENTIAL.
 FT TransmemEM 158 178 POTENTIAL.
 FT TransmemEM 192 212 POTENTIAL.
 FT TransmemEM 220 240 POTENTIAL.
 FT TransmemEM 260 280 POTENTIAL.
 FT TransmemEM 299 319 POTENTIAL.
 FT TransmemEM 319 337 POTENTIAL.
 FT TransmemEM 357 377 POTENTIAL.
 FT TransmemEM 391 411 POTENTIAL.
 FT TransmemEM 470 490 POTENTIAL.
 SQ SEQUENCE 496 AA; 52521 MW; OF1B23C8FD27BCDD CRC64;
 Query Match 62.2%; Score 28; DB 1; Length 496;
 Best Local Similarity 60.0%; Pred. No. 1e+02; 3; Indels 0; Gaps 0;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MGSIVLSQS 10
 DB 437 MGEAVQLANS 446
 RESULT 39
 CPJ5_MOUSE STANDARD; PRT; 501 AA.
 AC 054749; (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 40, Last sequence update)
 DE CYTOCHROME P450 2J5 (BC 1.14.14.1) (CYP11J5) (ARACHIDONIC ACID
 DE EPOXYGENASE).
 GN CYP2J5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC [11]TaxID=10090;
 RN [11]TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
 RX MEDLINE=98234557; PubMed=9570962;
 RA Ma J.; Ramchandran S.; Fiedorek F.T. Jr.; Zeldin D.C.;
 RA Mapping of the Cyt P450 cytochrome P450 genes to human chromosome 1 and
 RA Genomics 49:152-155(1998).
 RL Genomics 49:152-155(1998).
 CC -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
 CC -1- OXIDIZED FLAVOPROTEIN + H(2)O.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U62294; BAB87635.1;
 DR MGD: MGI:1270149; CYP2J5.
 DR InterPro: IPR001128;
 DR InterPro: IPR002401;
 DR Pfam: PF00067; p450; 1.
 DR PRINTS; PR00385; P450.

DR PRINTS; PR00463; EP450.
 DR PROSITE; PS00096; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
 KW Microsome; Endoplasmic reticulum.
 FT BINDING 447 447 HEME (BY SIMILARITY).
 SQ SEQUENCE 501 AA; 57784 MW; C67F2E79DD64AF99 CRC64;
 Query Match 62.2%; Score 28; DB 1; Length 501;
 Best Local Similarity 60.0%; Pred. No. 1e+02; 2; Indels 0; Gaps 0;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGSIVLSQS 10
 DB 371 MGNIVPLNS 380
 RESULT 40
 YHVO_YEAST STANDARD; PRT; 579 AA.
 AC P38848; 1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOPHETICAL 66.1 KDA PROTEIN IN IME3-SPO12 INTERGENIC REGION.
 GN YHR150W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OC [11]TaxID=4932;
 RN [11]TaxID=4932;
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 MEDLINE=94378003; PubMed=8091229;
 RA Johnston M.; Andrews S.; Brinkman R.; Cooper J.; Ding H.; Dover J.;
 Du Z.; Favell A.; Fulton L.; Gattung S.; Geisel C.; Kirsten J.;
 Kucaba T.; Hillier L.; Jier M.; Johnston L.; Lagosky J.; Mousier L.;
 Nian M.; Rifkin L.; Riles L.; St Peter H.; Travaskis E.; Vaughan K.;
 Vignati D.; Wilcox L.; Wohlman P.; Waterston R.; Wilson R.;
 Vaubin M.;
 RA *Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII.; 265,2077-2082(1994).
 RL Science 265:2077-2082(1994).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: U10397; AAB6880.1;
 DR EMBL: U10397; AAB6880.1;
 DR PIR: S46757; S46757.
 DR SGD: S0001193; YHR150W.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 397 421 POTENTIAL.
 FT TRANSMEM 397 421 POTENTIAL.
 SQ SEQUENCE 579 AA; 66148 MW; 2A6A73999C54C19E CRC64;
 Query Match 62.2%; Score 28; DB 1; Length 579;
 Best Local Similarity 60.0%; Pred. No. 1.2e+02; 3; Indels 0; Gaps 0;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MGSIVLSQS 10
 DB 219 MGSIFELQDS 228
 RESULT 41
 YEHQ_ECOLI

ID YEHQ_ECOLI STANDARD; PRT; 614 AA.
AC P3353;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 67.7 KDA PROTEIN IN MOLR-BGLX INTERGENIC REGION.
GN YEHQ.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
SC Science 277:1453-1474(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00007; AAA60485.1; ALT_INIT.
DR EMBL; AE000301; AAC75183.1; ALT_INIT.
DR Ecogen; EGI2003; yehq.
KW Hypothetical protein.
SQ SEQUENCE 614 AA; 67730 MW; 8056294BAE3CA56E CRC64;

Query Match 62.2%; Score 28; DB 1; Length 614;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 GSIVQLSQS 10
Db 365 GSILHLRS 373
|||: ||:|

RESULT 42
KDGLDROME STANDARD; PRT; 791 AA.
ID KDGLDROME
AC Q01583;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DIACYLGLYCEROL KINASE (EC 2.7.1.107) (DIGLYCERIDE KINASE 1) (DGK 1)
DE (DAG KINASE 1).
GN DGK OR DGK1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=92335231; PubMed=1321433;
RA Masai I., Hosoya T., Kojima S., Hotta Y.;
RT "Molecular cloning of a Drosophila diacylglycerol kinase gene that is

expressed in the nervous system and muscle.";
Proc. Natl. Acad. Sci. U.S.A. 89:6030-6034(1992).
[2]
RN SEQUENCE OF 279-791 FROM N.A.
RP STRAIN=CANTON-S;
RX MEDLINE=93143713; PubMed=8380995;
RA Harden N., Yap S.F., Chiam M.-A., Lim L.;
RT "A Drosophila gene encoding a protein with similarity to
RL diacylglycerol kinase is expressed in specific neurons.";
Biochem. J. 289:439-444(1993).
CC -!- FUNCTION: UPON CELL STIMULATION CONVERTS THE SECOND MESSENGER
CC DIACYLGLYCEROL INTO PHOSPHATIDATE, INITIATING THE RESYNTHESIS
CC OF PHOSPHATIDYLINOSITOLS AND ATTENUATING PROTEIN KINASE C
CC ACTIVITY (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + 1,2-DIACYLGLYCEROL = ADP +
CC 1,2-DIACYLGLYCEROL 3-PHOSPHATE
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE NERVOUS SYSTEM AND MUSCLE.
CC -!- DEVELOPMENTAL STAGE: DGK IS TRANSCRIBED IN THE EMBRYONIC, PUPAL
CC AND ADULT STAGES, WITH LITTLE EXPRESSION DURING THE LARVAL
CC STAGES. EXPRESSION IN LATE EMBRYOS IS SPECIFIC TO THE CENTRAL
CC NERVOUS SYSTEM AND HEAD.
CC -!- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D11120; BAA01894.1; -;
DR EMBL; X67335; CAA47750.1; -;
DR PIR; S25099; S25099.
DR PIR; S28229; S28229.
DR PIR; A46140; A46140.
DR FlyBase; FBgn0004568; Dgk.
DR InterPro; IPR000756; -;
DR InterPro; IPR001206; -;
DR Pfam; PF00609; DAGKa; 1.
DR Pfam; PF00781; DAGKc; 1.
KW Transferase; Kinase.
FT DOMAIN 344; 353
FT DOMAIN 401; 405
FT DOMAIN 430; 444
FT DOMAIN 534; 539
FT DOMAIN 130; 256
FT DOMAIN 564; 755
FT CONFLICT 374; 374
FT CONFLICT 444; 444
SQ SEQUENCE 791 AA; 87267 MW; 7D50D8ED01496679 CRC64;

Query Match 62.2%; Score 28; DB 1; Length 791;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 MGSIVQLSQS 10
Db 279 MQKVIQLSQS 288
|::|::|

RESULT 43
PERT_BORPE STANDARD; PRT; 910 AA.
ID PERT_BORPE
AC P14283;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PERTACTIN PRECURSOR (OUTER MEMBRANE PROTEIN P.69) (P.93).
GN PRN OR OMP69A.
OS Bordetella pertussis.

OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 CC Bordetella
 CC [1] TaxID=520;
 RN
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=ISOLATE CN2992;
 RX MEDLINE=9264462; PubMed=2542937;
 RA Charles I.G., Dougan G., Pickard D., Chatfield S., Smith M.,
 RA Novotny P., Morrissey P., Fairweather N.F.,
 RA Molecular cloning and characterization of protective outer membrane
 FT protein 66 kDa from *Bordetella pertussis*.
 RL Proc. Natl. Acad. Sci. U.S.A. 86:3554-3558(1989).
 RN
 RP REVISIONS TO 264 AND 332.
 RX MEDLINE=92407514; PubMed=1527510;
 RA Li J.L., Fairweather N.F., Novotny P., Dougan G., Charles I.G.,
 RA Cloning, nucleotide sequence and heterologous expression of the
 FT outer membrane protein P.68 pertactin from *Bordetella*
 RL *pertussis*.
 RN J. Gen. Microbiol. 138:1697-1705(1992).
 RP
 RX X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=96196517; PubMed=8609998;
 RA Embley P., Charles I.G., Fairweather N.F., Isaacs N.W.,
 RA Structural analysis of *Bordetella pertussis* virulence factor P.69 pertactin.
 RL Nature 381:90-92(1996).
 CC
 CC -1- FUNCTION: AGGLUTINOGEN THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
 CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
 CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
 CC
 CC -1- SUBUNIT: MONOMER.
 CC
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
 CC
 CC -1- DISEASE: PERTUSSIS.
 CC
 CC -1- CONCENTRATION: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
 CC
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL; J04560; AAA22980.1; ALT_SEQ.
 DR PIR; A32560; A32560.
 KW Outer membrane; signal; Virulence; Repeat.
 FT SIGNAL 1 34
 FT CHAIN 35 910
 FT PROPEP 35 910
 FT SITE 260 262
 FT
 FT DOMAIN 266 290
 FT REPEAT 266 270
 FT REPEAT 270 280
 FT REPEAT 280 285
 FT REPEAT 285 290
 FT REPEAT 290 293
 FT DOMAIN 579 593
 FT SEQUENCE 910 AA; 93452 MW; A169871E20A2E7DB CRC64;
 .SQ
 Query Match 62.2%; Score 28; DB 1; Length 910;
 Best Local Similarity 66.7%; Pred No. 2e+02; Indels 0; Gaps 0;
 Matches 6; Conservative 2; Mismatches 1;
 QY 2 GSVLSQS 10
 DB 302 GSSVELAQS 310
 |||1:1|||
 RESULT 44

TYCA_BACBR
 ID 10005
 AC 030407; STANDARD; PRT; 1088 AA.
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TYROCIDINE SYNTHETASE I [INCLUDES: ATP-DEPENDENT D-PHENYLALANINE
 DE ADENYLASE (D-PHEA) (D-PHENYLALANINE ACTIVASE); PHENYLALANINE RACEMASE
 DE [ATP-HYDROLYZING] (EC 5.1.1.11)].
 OS *Bacillus brevis*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC *Bacillus/Staphylococcus* group; *Brevibacillus*.
 OX NCBI_TaxID=1393;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9008354; PubMed=3267240;
 RA Weckermann R., Furbass R., Marahiel M.A.;
 RA "Complete nucleotide sequence of the tyca gene coding the tyrocidine
 FT synthetase 1 from *Bacillus brevis*.";
 RL Nucleic Acids Res. 16:11841-11841(1988).
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98012987; PubMed=9352938;
 RA Mootz H.D., Marahiel M.A.;
 RA "The tyrocidine biosynthesis operon of *Bacillus brevis*: complete
 FT nucleotide sequence and biochemical characterization of functional
 FT internal adenylation domains.";
 RL J. Bacteriol. 179:6843-6850(1997).
 RN
 RP SEQUENCE OF 1-62 FROM N.A.
 RX MEDLINE=87194610; PubMed=3032912;
 RA Marahiel M.A., Zuber P., Czekay G., Losick R.;
 RA "Identification of the promoter for a peptide antibiotic biosynthesis
 FT gene from *Bacillus brevis* and its regulation in *Bacillus subtilis*.";
 RL J. Bacteriol. 169:2215-2222(1987).
 CC
 CC -1- FUNCTION: THE FINAL STEP CATALYZES THE D-ISOMER
 CC CATALYZES PHENYLALANINE TO THE D-ISOMER.
 CC
 CC -1- CATALYTIC ACTIVITY: ATP + L-PHENYLALANINE -> AMP + PYROPHOSPHATE
 CC + D-PHENYLALANINE.
 CC
 CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETINE
 CC (POTENTIAL).
 CC
 CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
 CC
 CC -1- ANTI-BIOTIC TYROCIDINE BIOSYNTHESIS OF TYCA, TYCB AND TYCC.
 CC
 CC -1- SUBUNIT: LARGE MULTIMERIC COMPLEX OF TYCA, TYCB AND TYCC.
 CC
 CC -1- EPIMERIZATION DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO
 CC THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
 CC RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOLATION, CONDENSATION
 CC (NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
 CC N METHYLATION (OPTIONAL).
 CC
 CC -1- MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
 CC DECAPEPTIDES. TYROCIDINE IS A PHOSPHOPANTHETINE-GLN-TYR-VAL-
 CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
 CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
 CC
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE. 56% IDENTITY TO
 CC GRAMICIDIN S SYNTHETASE I (GRSA), WHICH CATALYZES THE SAME
 CC REACTION IN GRAMICIDIN S SYNTHESIS.
 CC
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 CC EMBL; X13237; CAA31623.1;
 DR EMBL; AF004835; AAC45928.1;

- Search completed: June 28, 2001, 11:57:31
- Job time: 272 sec